NASA TECHNICAL NOTE

NASA TN D-6073

NASA TN D-6073

DIGITAL ANALYSIS OF RANDOM DATA RECORDS BY PIECEWISE ACCUMULATION OF TIME AVERAGES

by F. R. Krause, J. A. Jones, M. J. Fisher, and J. C. Pooley

George C. Marshall Space Flight Center
Marshall Space Flight Center, Ala. 35812

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • DECEMBER 1970

1. Report No. NASA TN D-6073	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Digital Analysis of Random Data Records by Piecewise Accumulation of Time Averages		5. Report Date December 1970
		6. Performing Organization Code
7. Author(s) F. R. Krause, J. A. Jones, M. J. Fisher (IIT Research Institute - Now Rolls Royce Lecturer, University of Southampton, England), and J. C. Pooley (Nortronics-Huntsville, Northrop Corporation)		8. Performing Organization Report No.
9. Performing Organization Name and Address George C. Marshall Space Flight Center Marshall Space Flight Center, Alabama 35812		10. Work Unit No. M531
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address		13. Type of Report and Period Covered Technical Note
		14. Sponsoring Agency Code
15. Supplementary Notes Prepared by Aero-Astrodynamics and Computation Laboratories, Science and Engineering Directorate.		

16. Abstract

This report describes a statistical data reduction program designed to reduce and analyze random data processes without regard to record length. By this method, very long records may be processed even when computer storage limitations exist. Additionally, the statistical certainty is constantly checked and updated until the desired statistical accuracies are realized or the data source is exhausted. Consequently, only those data necessary to achieve the desired result are processed. This can achieve a considerable savings in computer time and computing costs.

17. Key Words (Suggested by Author(s))		18. Distribution Statement	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 64	22. Price* $ 3.00

*For sale by the National Technical Information Service, Springfield, Virginia 22151

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

DEFINITION OF SYMBOLS

Symbol	Definition
1. Independent Variables	
t	Observation time
T	Integration time
ΔT	Piece length
i	Piece number $(t/\Delta T)$
m	Accumulation number $(T/\Delta T)$
f	Frequency
τ	Time lag
τ_M	Maximum time lag
$k = 1, 2, \ldots N$	Number of imaginary realizations
2. Dependent Variables	
$x(t)$	Time history of record x
$y(t)$	Time history of record y
σ_i	Root mean square value of i
R	Product mean value of x and y
$C(t)$	Covariance function
ΔR	Statistical error of product mean value
δR	Averaged statistical error of product mean value

DEFINITION OF SYMBOLS (Continued)

Symbol	Definition
3. Operators	
$(\ )$	Statistic
$\overline{(\ )}_i = \frac{1}{\Delta T} \int_{(i-1)\Delta T}^{i\Delta T} (\ ) \, dt$	Piecewise mean
$\overline{\overline{(\ )}} = \frac{1}{m} \sum_{i=1}^{m} \overline{(\ )}_i$	Accumulative mean
$\overline{\Delta^2(\ )}_m = \frac{1}{m-1} \sum_{i=1}^{m} \left[\overline{(\ )}_i - \overline{(\ )} \right]^2$	Piecewise statistical error
$\Delta^2\overline{\overline{(\ )}}_m = \frac{\overline{\Delta^2(\ )}_m}{M}$	Accumulative statistical error
$E[\overline{(\ )}] = \frac{1}{N} \sum_{k=1}^{N} (\ )^{(k)}$	Sample of expected value or ensemble average for one group of N realizations
$\sigma^2 = (\sigma^{(1)}) = \frac{1}{N-1} \sum_{h=1}^{N} \left\{ \overline{(\ )}^{(h)} - E[\overline{(\ )}] \right\}^2$	Sample of variance between realizations
$_\tau\overline{(\ )} = \frac{1}{2\tau_{max}} \int_{-\tau_{max}}^{+\tau_{max}} \overline{(\ )} \, d\tau$	Average over time lags

DEFINITION OF SYMBOLS (Concluded)

SUBSCRIPTS

i	Time interval $(i-1)\Delta T \leqq t \leqq i\Delta T$
m	Accumulation over all pieces up to $i = m$
x	From record x
y	From record y

ACKNOWLEDGMENT

The authors wish to express their appreciation to Robert R. Jayroe, Jr., for his critical review and helpful suggestions concerning this paper. Our thanks also to Gerald J. Curet for his efforts in running the test cases through the data reduction processes.

DIGITAL ANALYSIS OF RANDOM DATA RECORDS BY PIECEWISE ACCUMULATION OF TIME AVERAGES

SUMMARY

A data reduction program for the reduction and analysis of random processes has been developed. The data are digitized, and subsequent statistical analyses are performed. The usual storage limitations concerning long data records are overcome by subdividing each record into truncated pieces and performing a statistical analysis of each piece. As each piece is analyzed, a recursive process is utilized to update average statistical parameters for all subsequent pieces until a predetermined statistical accuracy is realized or until all data have been exhausted.

The operator can exercise various control parameters and achieve desired statistical parameters, such as error analyses, correlation coefficient, power spectra density, transfer function, phase relationships, and coherence function. Single or multiple functions may be processed.

This program is coded in the compressed form of FORTRAN. This allows the operator to add and/or delete any specified portion of the existing program with alter cards.

INTRODUCTION

An experimental investigation of many complex physical phenomena, including atmospheric motions, random vibrations, turbulence, acoustics, and combustion processes requires the use of statistics for their quantification. This, in turn, leads to a requirement for methods of statistical data reduction which are accurate, convenient, and, if possible, versatile. With respect to versatility a digital computer program is particularly attractive because the modification of a reduction routine requires only the preparation of computer software as opposed to the often lengthy research and development process involved in the supply of analog equipment.

This report describes the development of a statistical data reduction routine in which digitization of analog data records is undertaken, subsequent statistical calculation being performed on a digital computer. No limitation exists regarding the length of records which may be processed. Such limitations are normally set by storage limitations of the machine. However, as demonstrated herein, this difficulty can be overcome by introducing the concept of "piecewise accumulated mean values."

A second important peripheral advantage of this concept is that it permits the statistical certainty of the result to be checked as the calculation proceeds. In fact, in the present program this check can be employed as a program control feature. The calculation is allowed to proceed only until a certain statistical certainty, specified by the user, is obtained. In this way no computer time is employed other than that which is necessary to reach the degree of statistical accuracy which the user feels to be necessary for his particular application. The discussion of these error estimates also shows how their variation as a function of record processed may be used in the detection of nonstationary trends in the record. Such trends are a common feature of investigations of uncontrolled environments, such as the atmosphere, and hence their reliable detection offers a valuable additional feature of the "piecewise" correlation approach.

The estimation of the first and second derivatives of covariance functions is discussed. Not only are these parameters frequently of value in their own right, but as demonstrated here they can be employed to interpolate the correlation function between measured lag points, thus providing an improved definition.

The report concludes with a description of tests, performed utilizing a band-limited white noise test signal, which demonstrate the features of the program outlined in the earlier sections.

ACCUMULATION OF PIECEWISE AVERAGES

The statistical descriptions of random processes are normally required in terms of mean values, root mean square values, correlation functions (and spectra), and probability statistics. The definition of these values as derived from a finite record of total length T may be written:

mean values:

$$\overline{X} = \frac{1}{T} \int_0^T x(t)\ dt \quad , \tag{1}$$

root mean square values:

$$\sigma = \left[\frac{1}{T} \int_0^T x(t)^2\ dt \right]^{1/2} \quad , \tag{1.1}$$

correlation functions:

$$\overline{R(\tau)} = \frac{1}{T} \int_0^T x(t)\ y\ (t + \tau)\ dt \quad , \tag{2}$$

probability statistics:

The n^{th} moment of the probability distribution is

$$\overline{x^n} = \frac{1}{T} \int_0^T x^n(t)\ dt \quad . \tag{3}$$

We may generalize this problem, for convenience of the present discussion, by the schematic representation:

$$\overline{(\quad)} = \frac{1}{T} \int_0^T (\quad)\ dt \quad , \tag{4}$$

where we refer to the ( ) operations as a statistic and note that they normally involve some multiplicative operation as indicated by equations (2) and (3).

The normal problem which arises in programming our generalized equation (4) for computer operation is that of providing sufficient storage locations for the basic data $x(t)$ or $y(t)$ if the integration period T is to be of sufficient length to yield a statistically significant result. For example, in terms of the test cases discussed at the end of this report, a total record of 30-sec duration has been processed with an analog-to-digital conversion rate of 20 000 sample pairs per second. Thus, 1.2×10^6 storage locations would have been needed for storage of data alone, irrespective of the requirements for the executive portion of the program.

To avoid such a problem we subdivide the total record into m equal pieces each of length T and introduce data from each piece into the machine separately (Fig. 1). The piecewise average of a statistic is then calculated by integrating initially over each one of these pieces. This leads to the definition of a piecewise or sample mean, which for the i^{th} piece is

$$\overline{(\ )}_i = \frac{1}{\Delta T} \int_{t=(i-1)\Delta T}^{t=i\Delta T} (\ )\ dt \quad . \tag{5}$$

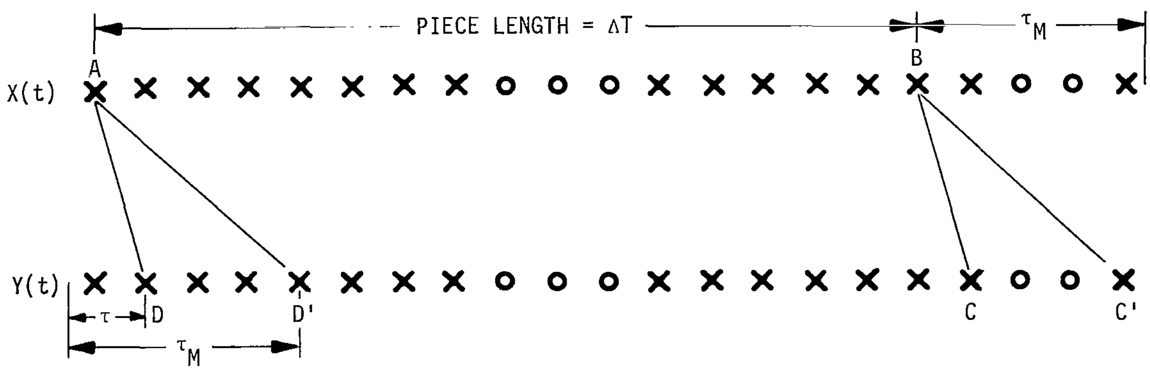


Figure 1. Subdivided records.

The desired average over the entire record is subsequently obtained by averaging the result obtained from all pieces. This leads to the definition of an accumulative mean,

$$\overline{\overline{(\ )}}_m = \frac{1}{m} \sum_{i=1}^{m} \overline{(\ )}_i \quad . \tag{6}$$

Expanding the summation we find that

$$\overline{\overline{(\ )}}_m = \frac{1}{m}\frac{1}{\Delta T}\int_0^{\Delta T}(\ )\,dt + \ldots \frac{1}{\Delta T}\int_{(i-1)\Delta T}^{i\Delta T}(\ )\,dt + \ldots \frac{1}{\Delta T}\int_{(m-1)\Delta T}^{m\Delta T}(\ )\,dt$$

$$= \frac{1}{T}\int_0^T (\ )\,dt \quad . \tag{7}$$

The accumulative mean is thus identified with the time integral over the entire record as long as the integrand, $(\ )_i$, may be derived directly from the signals actually recorded. In the succeeding section we discuss the problem of errors which arise when this is not the case. However, although this method of approach has appreciably reduced the data storage requirements, equation (6) would still indicate the necessity of storing the m sample or piecewise means. This, also, can be avoided by the use of the following recursion formula. Suppose we obtained an accumulative mean over the first m pieced and have just processed a new piece $(m + 1)$. The mean over all pieces including the new one,

$$\overline{\overline{(\ )}}_{m+1} = \frac{1}{m+1}\sum_{i=1}^{m+1}\overline{(\ )}_i$$

$$= \frac{m}{m+1}\left\{\frac{1}{m}\sum_{i=1}^{m}\overline{(\ )}_i + \frac{1}{m}\overline{(\ )}_{m+1}\right\} \quad , \tag{8}$$

can then be found by updating the information of the previous m pieces by including the mean $(\ )$ of the new piece. Substitution of equation (6) into equation (8) leads to the recursion formula:

$$\overline{\overline{(\ )}}_{m+1} = \frac{m}{m+1}\overline{\overline{(\ )}}_m + \frac{1}{m+1}\overline{(\ )}_{m+1} \quad . \tag{9}$$

Therefore, the need to store the m individual piecewise averages is eliminated. Only the accumulative mean of all pieces processed to date is required and can be updated each time a new piecewise mean is obtained. This effectively reduces all limitations on the length of record which can be processed.

A second advantage is that the latest estimate of the mean accumulated to date is always available for output and inspection. This facility has proved extremely valuable in a number of diagnostic applications.

Subdivision Error

As pointed out, the procedure leading to the recursion formula in equation (9) yields a result identical with the time averaged mean only if the multiplicative process used in forming the statistic involves only the actually recorded or digitized data values. A prime example, in point, where this is not the case arises in the measurement of the covariance of two signals having nonzero mean values. The time-averaged value of the covariance is given by

$$C_T\ (\tau=0) = \frac{1}{T} \int_0^T [x(t) - \bar{\bar{x}}_m]\ [y(t) - \bar{\bar{y}}_m]\ dt \tag{10}$$

$$= \frac{1}{T} \int_0^T x(t)\ y(t)\ dt - \bar{\bar{x}}_m \bar{\bar{y}}_m \tag{11}$$

The piecewise estimate of this covariance is based, however, on the contributions from single pieces, which consider only the mean value of that piece.

The contribution to the covariance of the i^{th} piece is therefore

$$C_i\ (\tau=0) = \frac{1}{\Delta T} \int_{t=(i-1)\Delta T}^{i\Delta T} [x(t) - \bar{x}_i]\ [y(t) - \bar{y}_i]\ dt\ . \tag{12}$$

Applying the summation equation (6) to these contributions, we find that the accumulative covariance is:

$$C_m(\tau=0) = \frac{1}{m} \sum_{i=1}^{m} C_i(\tau=0)$$

$$= \frac{1}{m\Delta T} \sum_{i=1}^{m} \int_{(i-1)\Delta T}^{i\Delta T} [x(t) - \overline{x}_i]\ [y(t) - \overline{y}_i]\ dt$$

$$= \frac{1}{m\Delta T} \int_{0}^{m\Delta T} x(t)\ y(t)\ dt + \frac{1}{m} \sum_{i=1}^{m} \left[\overline{x}_i \overline{y}_i - \frac{\overline{x}_i}{\Delta T} \int_{(i-1)\Delta T}^{iT} y(t)\ dt - \frac{\overline{y}_i}{\Delta T} \int_{(i-1)\Delta T}^{i\Delta T} x(t)\ dt \right] \quad . \tag{13}$$

The first and third terms in the [ ] of equation (13) cancel, and the accumulative covariance becomes

$$\overline{\overline{C}}_m(\tau=0) = \frac{1}{T} \int_{0}^{T} x(t)\ y(t)\ dt - \frac{1}{m} \sum_{i=1}^{m} \overline{x}_i \overline{y}_i \quad . \tag{14}$$

This accumulative covariance thus differs from the time average, equation (11), by the subdivision error $S(m; \tau=0)$ where

$$S(m;\tau=0) = C_T(\tau=0) - C_m(\tau=0) = -\overline{\overline{x}}_m \overline{\overline{y}}_m + \frac{1}{m} \sum_{i=1}^{m} \overline{x}_i \overline{y}_i \quad ,$$

which can be rewritten in the form,

$$S(m;\tau=0) = \frac{1}{m} \sum_{i=1}^{m} (\overline{x}_i - \overline{\overline{x}}_m)\ (\overline{y}_i - \overline{\overline{y}}_m) \quad . \tag{15}$$

The value of this error in two limiting cases is quickly established. First, in the limit $m=1$ $(\Delta T=T)$ the error is zero by definition. In the opposing limit $m \rightarrow \infty$ we see that in practice this would mean that each piece contained only one data sample; thus, for each piece,

$$x(t) = \overline{x}_i \text{ and } y(t) = \overline{y}_i \quad .$$

Thus, referring to equation (12) we see that each $C_i(\tau=0)$ will be identically equal to zero. Thus, as $m \to \infty$,

$$\overline{\overline{C}}_m(\tau=0) \to 0 \quad .$$

For intermediate piece lengths $0 < \Delta T < T$ an idea of the significance of the subdivision error can be gained by consideration of a simple analog. Suppose the values of $(\overline{x}_i - \overline{\overline{x}}_m)$ and $(\overline{y}_i - \overline{\overline{y}}_m)$, respectively, were plotted as a function of i or its equivalent time through the record. Two "signals" would result subject to the condition that their minimum time period must be comparable to the piece length. Furthermore, equation (15) suggests that the subdivision error is determined by the covariance of our two artificial signals. Thus we reach the conclusion that a finite piece length acts like a high-pass filter which cuts off the energy contained in frequencies whose period is comparable to or greater than the piece length.

Tests described in a later section of this report confirm this hypothesis. It is shown that the indicated energy of the signal does begin to decrease once the piece length becomes comparable to the period of energy containing frequencies. However, the filtering action is very weak, being comparable to that of a high-pass filter having a cutoff frequency f_c given by

$$f_c = \frac{1}{\Delta T}$$

and a "rolloff" rate of only 2 to 3 dB per octave.

It is appropriate to conclude this section by pointing out that subdivision error is not a basic feature of piecewise operation per se. It arises only because of the particular method adopted here in which the piecewise mean values $\overline{x}_i$ and $\overline{y}_i$ are employed in the covariance estimates, equation (12). We have already noted that no subdivision error results as long as the actually recorded or digitized values are used in unmodified form through the accumulative procedures. Thus, referring to equation (11) we see that the first term of the root mean square of this expression could be accumulated without error and the product $\overline{x}_m \, \overline{y}_m$ subtracted at the end of the calculation when the necessary information was available. However, in the present program we believed that the filtering action of finite piece lengths did offer a feature of potential value for the elimination of slow drifts or trends when such events were present and undesirable. To eliminate subdivision error it is only necessary to choose piece lengths comparable to the period of the minimum frequencies of interest. Furthermore, experience has shown that this is seldom the consideration which controls minimum allowable piece length, particularly where cross-correlation analysis is required. However, a general awareness of this feature is important.

ENVIRONMENTAL AND STATISTICAL VARIATION OF PIECEWISE AVERAGES

As mentioned in the introduction to this report, piecewise accumulation of mean values offers the opportunity to estimate the statistical reliability of a result as the calculation proceeds. In fact, in the existing program this parameter is used in formulating a decision as to whether another piece of data should be admitted. This section of the report outlines this approach, and test cases which appear to justify its use are presented in the section, Demonstration of the Piecewise Method.

The value of any piecewise average $\overline{(\ )}_i$ will depend both on its position in the complete record and on the piece length ΔT. These variations of piecewise values may be caused either by a change of the environment (i.e., a nonstationarity) or by a lack of integration time. In any finite integration period a certain amount of statistical variation is to be expected since only a finite number of random occurrences can take place, and these will not all cancel precisely.

Unfortunately, environmental and statistical errors occur simultaneously and are difficult to separate. Separation may, however, be achieved in theory by regarding the actual conducted experiment as one sample of a population of imaginary experiments which are all recorded for identical time-independent boundary conditions. Assume that $j = 1, 2, 3, \ldots N$ realizations of the same environment have been observed. Statistical averages may then be established by averaging over the different realizations instead of over time. This ensemble average will be denoted by $E[(\ )]_N$ and termed the expected value:

$$E[(\ )]_N = \frac{1}{N} \sum_{j=1}^{N} (\ )^{(j)} \quad .$$

Let $x^{(j)}(t)$ and $y^{(j)}(t)$ denote two signal records from the j^{th} realization.

The contribution to the correlation obtained from this j^{th} piece of length ΔT is

$$\overline{R}_j = \frac{1}{\Delta T} \int_{t_j}^{t_j + \Delta T} (\ )^j \, dt \quad , \tag{16}$$

while the ensemble average is obtained by summing over all N available realizations:

$$E(\overline{R}_i)_N = \frac{1}{N} \sum_{i=1}^{N} \overline{R}_i \quad . \tag{17}$$

Each sample mean R_i will differ from the ensemble average, $E(R_i)$, because of statistical or other variations. A measure of the magnitude of these variations of sample means is provided by the mean square error or variance:

$$(\mathrm{Var}\ R_i)_N = \frac{1}{N-1} \sum_{i=1}^{N} \left[R_i - E(R_i)_N \right]^2 \quad . \tag{18}$$

The practical implications of the central limit theorem now imply that variations of the piecewise time averages relative to their population mean should be normally distributed, regardless of the physical process generating the time history record. One estimate of this population mean is provided by the ensemble average, $E(R_i)_N$. Further, the probability that any sample mean will deviate from the ensemble average by more than a prescribed amount is given by the student "t" distribution. We may thus calculate a certain confidence limit for the variation of the sample means around the ensemble average. This limit will not be exceeded by a fraction, p, of all realizations. The normalized value of the confidence level, t_p, is called the confidence factor and is expressed in the inequality

$$|R_i - E(R_i)| \leq t_p(N) \ (\mathrm{VAR}\ \overline{R}_i)_N^{1/2} \quad . \tag{19}$$

The confidence factors for the normal distribution are listed in Table 1.

TABLE 1. PERCENTILE FACTORS AND CONFIDENCE INTERVALS FOR STATIONARY RECORDS

Student's t Distribution			χ^2-Distribution			
m	$t_{0.90}$	$1/\sqrt{m}$	$\chi^2_{0.50}$	$\chi^2_{0.10}$	$\chi_{0.90}/m$	$\chi_{0.10}/m$
2	3.08	0.707	2.71	0.0158	0.823	0.063
3	1.89	0.578	4.61	0.211	0.716	0.153
4	1.64	0.500	6.25	0.584	0.625	0.191
5	1.53	0.447	7.78	1.06	0.558	0.206
6	1.45	0.408	9.24	1.61	0.507	0.211
7	1.44	0.378	10.6	2.20	0.465	0.212
8	1.42	0.353	12.0	2.83	0.433	0.210
9	1.40	0.333	13.4	3.49	0.407	0.208
10	1.38	0.317	14.7	4.17	0.383	0.204
12	1.35	0.288	17.3	5.58	0.347	0.201
14	1.34	0.267	19.8	7.04	0.318	0.190
16	1.33	0.250	22.3	8.55	0.295	0.183
18	1.33	0.236	24.8	10.1	0.277	0.179
20	1.32	0.222	27.2	11.7	0.261	0.174
25	1.32	0.200	33.2	15.7	0.230	0.160
30	1.31	0.183	39.1	19.8	0.028	0.150
61	1.30	0.128	74.4	46.5	0.142	0.112
$m \to \infty$	1.28	$1/\sqrt{m}$	m	m	$1/\sqrt{m}$	$1/\sqrt{m}$

Equation (18) does give an estimate for the variations of the individual realizations, and the associated confidence intervals could be calculated if $(\mathrm{Var}\ R_i)_N$ were known accurately. However, we have calculated $(\mathrm{Var}\ R_i)_N$ from a finite number of realizations, N, of the experiment. Hence, in another series of N realizations a different value of this variance might be obtained. An accurate description of the statistical variations should therefore include not only the variations between individual realizations within a single group but also variations between different groups of realizations.

Let Var $\overline{R}_i$ denote the population variance which is calculated by taking the arithmetic mean of all samples of $(\mathrm{Var}\ R_i)_N$, i.e., the mean variance for all groups. The relative variation between the variance estimate from a single group and the average over all groups can then be expressed by the new variable:

$$\chi^2 = \frac{N(\mathrm{Var}\ \overline{R}_i)_N}{\mathrm{Var}\ \overline{R}_i} \quad . \tag{20}$$

The probability distribution of this variable is given by another universal distribution function, the χ^2 distribution. Knowing this distribution, one can calculate a lower limit, $\chi^2_{0.10}(N)$, which will be exceeded by the χ^2 samples of all but 10 percent of the admitted groups. One can also calculate an upper limit $\chi^2_{0.90}(N)$, which will exceed 90 percent of all χ^2 samples. Both limits together then give a confidence interval for the statistical variation of variance estimates between different groups of realizations. The 80-percent confidence interval would be

$$\frac{N(\mathrm{Var}\ \overline{R}_i)_N}{\chi^2_{0.90}(N)} \leqq \mathrm{Var}\ \overline{R}_i \leqq \frac{N(\mathrm{Var}\ \overline{R}_i)_N}{\chi^2_{0.10}(N)} \quad . \tag{21}$$

The confidence factors $\chi^2_{0.90}$ and $\chi^2_{0.10}$ are also listed in Table 1. For $N \geqq 30$ they may be calculated from the confidence factors, Z_p, of the normal distribution [1]:

$$\chi_p^2 = \frac{1}{2}\left(Z_p + \sqrt{2(N-1)} - 1\right)^2 = N\left(1 - \frac{3}{2N}\right)\left(1 + \frac{Z_p}{\sqrt{2N-3}}\right)^2 . \tag{22}$$

In particular, $Z_{0.90} = 1.28$, and $Z_{0.10} = 1/1.28 = 0.78$.

In summary, we find that a single group of N realizations can provide the following estimates:

1. The ensemble average, $E(R_i)_N$, from equation (17).

2. The sample variance, $(\mathrm{Var}\ R_i)_N$, for the required statistical variation of sample means relative to the ensemble average from equation (18).

3. The confidence interval for the statistical variations of the sample variance relative to the population variance of many such sample groups.

All these estimates are based on universal distribution functions which are independent of the particular physical process from which the recorded time history is obtained.

The preceding study of statistical variations between piecewise means requires many samples which have been acquired by repeating the same experiment many times under the same environmental conditions. Such control of the environment is not always possible during space flight or in meteorological experiments.

The alternative is then to assume that the environmental boundary conditions are sufficiently time invariant during one experiment so that individual pieces of a long record represent statistically independent realizations of these invariant boundary conditions. For this purpose a long record of length T is subdivided into $i = 1, 2, 3, \ldots\ldots, m$ pieces of length $\Delta T = \frac{T}{m}$, as discussed in the section, Accumulation of Piecewise Averages. Each of these piecewise estimates is then treated as if it came from a new realization. This means that the summation over realizations is replaced by an accumulation of pieces. The estimate of the ensemble average for any statistic $(\ )_i$ becomes

$$E[\overline{(\ )}_i]_N = \frac{1}{N}\sum_{j=1}^{N}\overline{(\ )}_i^{(j)} \Rightarrow \frac{1}{m}\sum_{i=1}^{m}\overline{(\ )}_i^{\,j} = \overline{\overline{(\ )}}_m \quad . \tag{23}$$

Similarly, the sample variance becomes

$$[\mathrm{Var}\,\overline{(\ )}_i]_N = \frac{1}{N-1}\sum_{j=1}^{N}\left\{(\ )_i^j - E[\overline{(\ )}_i]_N\right\}^2$$

$$\Rightarrow \frac{1}{m-1}\sum_{i=1}^{m}\left[\overline{(\ )}_i - \overline{\overline{(\ )}}_m\right]^2 = \Delta^2\overline{(\ )}_m \quad . \tag{24}$$

The following conditions [2] must be met to justify this replacement of realizations with pieces:

1. The time history of the statistic "( )" is a self-stationary process.

2. The autocovariance function of this time history meets certain integrability conditions.

3. The individual piece lengths, ΔT, exceed the time-lag range within which the autocovariance is significant.

Experimental data usually can be arranged to meet the preceding conditions 2 and 3. However, condition 1 requires that the environmental variations be negligible during the entire processed record. Such time historics are referred to as stationary. For such stationary time series all remaining variations are statistical. Thus variations between piecewise averages and the statistical variations between sets of pieces should all follow the universal probability distributions just discussed. The "fit" of these distributions may thus be used as a criterion for stationarity. One such criterion is developed in the remainder of this section.

Consider the confidence interval for the statistical variations of the sample variance $\Delta^2\overline{(\ )}_m$ of a piecewise average. This is obtained by combining equations (21) and (24) and yields

$$\frac{m\Delta^2\overline{(\ )}_m}{\chi^2_{0.90}(m)} \leqq \mathrm{Var}\,\overline{(\ )} \leqq \frac{m\Delta^2\overline{(\ )}_m}{\chi^2_{0.10}(m)} \quad . \tag{25}$$

However, the two factors $\frac{m}{\chi^2_{0.90}}$ and $\frac{m}{\chi^2_{0.10}}$ both asymptotically approach unity [equation (22)]. Thus in this limit the inequality of equation (25) approaches an equality yielding the population variance. The experimentally accessible statistical error of piecewise mean values will therefore approach a finite constant value if the environment is stationary.

The statistical error of a piece depends on piece length, and thus its value is rather arbitrary. One would expect it to decrease with increasing piece length since more statistical variations are included. Thus it would be a minimum if the piece spanned the entire available record. However, in this case no estimate of a variance could be made since this requires several measurements. If the process is stationary, the piecewise estimates are normally distributed and statistically independent, and the variance of the accumulative average from m pieces will be 1/m times that of the individual pieces.

Thus we obtain the desired expression for variance of the accumulated means:

$$[\mathrm{Var}\,\overline{\overline{(\ )}}_m]_N = \left[\mathrm{Var}\,\frac{1}{m}\sum_{i=1}^{m}\overline{(\ )}_i\right]_N = \frac{1}{m}\,[\mathrm{Var}\,\overline{(\ )}_i]_N \quad . \tag{26}$$

Replacing the variances with the experimentally accessible approximations contained in equation (24), we obtain

$$[\Delta\overline{\overline{(\ )}}_m]^2 = \frac{1}{m}\,[\Delta\overline{(\ )}_m]^2 = \frac{1}{m(m-1)}\sum_{i=1}^{m}\left[\overline{(\ )}_i - \overline{\overline{(\ )}}_m\right]^2 \quad . \tag{27}$$

Note that this is just the quantity required since it represents an estimate of the variance of a number of identical experiments, of which the one actually conducted is one member. That is, we have obtained the probable deviation of the mean actually obtained from a mean obtained from an ensemble of such experiments.

Substituting $m = T/\Delta T$ we find the well-known result that the probable error of an accumulative mean should decrease with the inverse square root of the integration time or record length, i.e.,

$$\Delta\overline{\overline{(\ )}}_m = \frac{\sqrt{\mathrm{Var}\,\overline{(\ )}_i\,\Delta T}}{\sqrt{T}} \quad . \tag{28}$$

The full practical significance of the error analysis of this section can now be explained by comparing the situation developed here with that existing where a pure time integration has been undertaken. In the latter case the ratio of the probable error of the mean value to the mean value is commonly expressed as

$$\frac{\Delta\sigma}{\sigma} = \frac{1}{\sqrt{BT}} \quad , \tag{29}$$

where B is the bandwidth of the data to be processed. However, this bandwidth is not known a priori, and at best some estimate based on a past experience or pure guess must be inserted in deciding on an integration time. At the end of this integration time, no method is available to ascertain whether the expected confidence level was reached, whereas any nonstationary trends in the data will have been observed in the integration process.

By contrast, using a piecewise method the uncertainty of the result obtained after processing any number of pieces can be determined by using equation (27), combined, if necessary, with confidence factors just discussed. Once the desired confidence interval is reached, the calculation can be terminated. No arbitrary decisions are required.

Finally, using equation (28) we see that $\Delta\overline{\overline{(\ )}}_m$ should decrease as the reciprocal square root of integration time or number of (equal length) pieces processed to date.

Of course, a precise straight-line relationship between $\Delta\overline{\overline{(\ )}}_m$ and $1/\sqrt{BT}$ is not to be expected since the values of $\Delta\overline{\overline{(\ )}}_m$ are subject to some uncertainty. However, the shape of the best straight-line fit can be employed to obtain an estimate of the equivalent bandwidth B. Finally, this bandwidth can be used to calculate confidence levels for the relative accumulative error.

Rearranging the two inequalities of equation (25) and using equations (25) and (28), we find the limits of uncertainty of the accumulative error specified by

$$\frac{\chi_{0.10}(m)\sqrt{m}}{\sqrt{BT}} \leqq \frac{\overline{\overline{(\;)}}_m}{|\overline{\overline{(\;)}}_m|} \leqq \frac{\chi_{0.90}(m)\sqrt{m}}{\sqrt{BT}} \quad . \tag{30}$$

Thus these limits can be marked around the best straight-line fit of the $\Delta\overline{\overline{(\;)}}_m$ vs $1/\sqrt{T}$ graph. If the majority of values are found contained within the intervals so specified, stationarity of the data record has been demonstrated. On the other hand, significant and consistent trends of values outside these limits indicate that a nonstationarity has occurred. Furthermore, the integration time for which such deviations occur indicates the onset of the nonstationarity and can be used to optimize an available record.

A direct calculation of the accumulative statistical error for the autocorrelation function of a band-limited white noise test signal is discussed in the section, Demonstration of the Piecewise Method. It can be seen that the predicted straight-line relationship is obeyed while the slope of the line, 700 Hz, equals the bandwidth of the filter used.

Also, the 90-percent confidence intervals calculated from consideration of the χ^2 distribution are demonstrated. The fact that all accumulative errors fall within the interval indicates the stationarity of the test signal.

In summary, the purpose of this section has been to demonstrate the way in which piecewise accumulation of statistical values can be employed to permit direct calculation of the probable error of the final result. No arbitrary decisions with regard to integration time are required. Calculation is allowed to proceed only until the required confidence interval is obtained. Furthermore, consideration of the behavior of these confidence intervals as a function of the length of processed record may be used to detect any nonstationary trends contained within that record.

COVARIANCE FUNCTIONS AND NOISE

The covariance function for two time history records, $x(t)$ and $y(t)$, respectively, is defined.

$$C_T(\tau) = \frac{1}{T} \int_0^T [x(t) - \bar{\bar{x}}_m] [y(t+\tau) - \bar{\bar{y}}_m] \, dt \quad . \tag{31}$$

Two problems arise in the programming of this algorithm for piecewise operation. First, as in the case of the zero time delay covariance $C_T(\tau=0)$ calculation discussed in the section, Accumulation of Piecewise Averages, the mean values of the signals $\bar{\bar{x}}_m$ and $\bar{\bar{y}}_m$ are not available during the accumulation procedure — only at its termination. As in the previous case they are replaced by the more accessible piecewise mean values $\bar{x}_i$ and $\bar{y}_i$, respectively. As was the case in the section just mentioned, this again leads to a subdivision error. Arguments similar to those previously used show that this error is eliminated as long as the piece length is comparable to or longer than the period of the minimum frequency of interest.

The second problem arises for the finite time delay values. In moving along the samples representing $x(t)$, one eventually arrives at the situation where the required $y(t+\tau)$ sample is not contained within the piece currently in the machine. The method used to correct this and maintain the same number of samples for each time delay is to read in an additional maximum time delay (τ_M) data sample. As the time delay (τ) shifts, τ additional samples are added to the delayed channel, maintaining the piece length. By this method the same statistical accuracy of the correlation is maintained for each time delay.

The situation is shown schematically in Figure 1. For the case illustrated in that figure the piece length is five times τ_M. Parallelogram ABCD shows the sample products for τ and parallelogram ABC'D' for τ'.

The piecewise covariance for time delay τ obtained from the i^{th} piece therefore is evaluated by the following algorithm:

$$\overline{C}_i(\tau) = \frac{1}{\Delta T} \int_{(i-1)\Delta T}^{i\Delta T} [x(t) - \overline{x}_i] \, [y(t+\tau) - \overline{y}_i] \, dt \quad . \tag{32}$$

Covariance functions are frequently used to determine which components of the signals contained in time records $x(t)$ and $y(t)$ are common to both. This concept of common signal is based on the time history of the instantaneous product, $x(t)\,y(t+\tau)$. The mean value of such a product is called the correlation function.

In accordance with our previous discussion, its piecewise average would be

$$\overline{R}_i(\tau) = \frac{1}{\Delta T} \int_{(i-1)\Delta T}^{i\Delta T} x(t)\, y(t+\tau)\, dt \quad . \tag{33}$$

It differs from the piecewise estimated covariance by the product of the piecewise mean values. Comparing equations (32) and (33) we find that

$$\overline{C}_i(\tau) = R_i(\tau) - \overline{x}_i\, \overline{y}_i \quad . \tag{34}$$

The instantaneous product, $x(t)\,y(t+\tau)$, will oscillate around its mean value, $R_i(\tau)$, and these oscillations will tend to cancel one another, at least partially, when the product is averaged by integrating over time.

Two signals are said to have no common modulations if their covariance vanishes. This complete cancellation will occur if, for example, an increase in one signal is accompanied by either an increase or decrease of the other, the two being equally probable. A typical example of two such signals is represented by shot noise emission from two photodetectors.

The partial cancellation of the oscillations of an instantaneous product provides a possible definition for noise components of a signal. The noise component, $x_N(t)$, of the total signal, $x(t)$, may be defined with respect to the second signal $y(t)$. It is that component of $x(t)$ which does not contribute to the covariance of $x(t)$ and $y(t)$, i.e.,

$$\overline{[x_N(t)\ y(t+\tau)]_i} = 0 \quad .$$

Conversely, the common component

$$x_c = x - x_N$$

is that component of signal $x(t)$ which is responsible for the finite value of the product mean value.

$$\overline{[x(t)\ y(t+\tau)]_i} = \overline{[(x_c + x_N)\ y(t+\tau)]_i} = \overline{[x_c(t)\ y(t+\tau)]_i} \quad . \tag{35}$$

The second signal $y(t)$ can be similarly divided into its common and noise component by now taking $x(t)$ as the reference signal, i.e.,

$$y(t) = y_c(t) + y_N(t) \quad . \tag{36}$$

Combining equations (35) and (36) we find that

$$\overline{[x(t)\ y(t+\tau)]_i} = \overline{[x_c(t)\ y_c(t+\tau)]_i} \tag{37}$$

The integration of the instantaneous product has thus eliminated the noise components, and the resultant product mean value depends only on the signal variations which are common to both signals. We should note here, however, that the word "common" does not imply that x_c and y_c are identical, merely that they have a fixed time invariant phase relationship. Furthermore, this phase relationship and their frequency content will determine the value of τ for which equation (37) obtains a maximum value.

For any finite piece length the noise components will not cancel precisely and will therefore produce statistical variations of the product mean values. One expects, therefore, that a relationship exists between the statistical variations of the piecewise means and the spectral distribution of both the common signal and noise signal power.

Jayroe and Su [3] have derived this relation analytically on terms of the cross-power spectral density, $G(f)$, of the common signals and the power spectral densities, $S_N(f)$, of the noise components. Their result may be stated as:

$$\Delta^2 \bar{\bar{C}}_m(\tau, T) = \frac{1}{m(m-1)} \sum_{i=1}^{m} (\bar{C}_i - \bar{\bar{C}}_m)^2$$

$$= \frac{1}{T} \int_{-f_2}^{+f_2} [G^2(f) + S^2_{Nx}(f) + S^2_{Ny}(f)] \, (1 + e^{-4\pi i f \tau}) \, df \quad . \qquad (38)$$

The power spectral densities are directly related to the mean square amplitudes of the associated signal components by:

$$\overline{\overline{(x_c - \bar{\bar{x}}_m)\,(y_c - \bar{\bar{y}}_m)}} = \int_{-f_2}^{+f_2} G(f)\, df \quad , \qquad (39)$$

$$\overline{\overline{(x_N - \bar{\bar{x}}_m)_m^{\,2}}} = \int_{f_1}^{f_2} S_{Nx}(f)\, df \quad , \qquad (40)$$

and

$$\overline{\overline{(y_N - \bar{\bar{y}}_m)_m^{\,2}}} = \int_{f_1}^{f_2} S_{Ny}(f)\, df \quad . \qquad (41)$$

Assuming that the noise components have a flat spectrum,

$$S_{Nx}(f) = \left| \begin{array}{ll} \dfrac{\overline{\overline{(X_N - \bar{\bar{X}}_m)_m^{\,2}}}}{f_2 - f_1} & \text{for } f_1 \leqq f \leqq f_2 \\ 0 \quad \text{elsewhere ,} & \end{array} \right. \qquad (42)$$

and

$$S_{Ny}(f) = \begin{cases} \dfrac{\overline{\overline{(y_N - \overline{\overline{y}}_m)_m^2}}}{f_2 - f_1} & \text{for } f_1 \leqq f \leqq f_2 \\ 0 \quad \text{elsewhere} , \end{cases} \tag{42.1}$$

the above relation may be rewritten as

$$\Delta^2 \overline{\overline{C}}_m(\tau_1 T) = \frac{1}{T} \int_{-f_2}^{+f_2} G^2(f)\ (1 + e^{-4\pi\, i f \tau})\ df \tag{42.2}$$
$$+ \frac{\overline{\overline{(x_N - \overline{\overline{x}}_m)_m^2}} + \overline{\overline{(y_N - \overline{\overline{y}}_m)_m^2}}}{(f_2 - f_1)\, T} + \frac{\sin 4\pi f_2 \tau - \sin 4\pi f_1 \tau}{4\pi (f_2 - f_1)\tau} .$$

This relation indicates that the statistical error accounts for the fluctuations of both the common signals and the noise components. Both contributions vanish with the inverse square root of the integration time, and both depend on the time lag in a known fashion. The time-lag dependence may thus be removed by integration. Integrating the time-lag-dependent factor in the integrand leads to an approximation of the Dirac function δ (f). The integral of the time-lag-dependent factor of the noise term vanishes in good approximation, since the factor oscillates between positive and negative values. The correlation of the time-lag dependence by integration thus gives the approximation

$$\tau\, \Delta^2 \overline{\overline{C}}_m = \frac{1}{2\tau_m} \int_{-\infty}^{+\infty} \Delta^2\, \overline{\overline{C}}_m(\tau, T)\ d\tau$$

$$= \frac{1}{T} \int_{-f_2}^{+f_2} G^2(f)\ df + \frac{G^2(f=0)}{T} \tag{43}$$
$$+ \frac{\left[\overline{\overline{(x_N - \overline{\overline{x}}_m)_m^2}}\right]^2 + \left[\overline{\overline{(y_N - \overline{\overline{y}}_m)_m^2}}\right]^2}{(f_2 - f_1)\ T} .$$

The second term characterizes the power at two frequencies and thus accounts for near-zero frequency components.

For autocorrelations ($y=x$) the noise terms vanish by definition. For our test case of band-limited white noise, the cross-power spectrum becomes the power spectrum.

$$G(f) = \begin{cases} \bar{\bar{C}}_m(0) / (f_2 - f_1) & f_1 \leqq f \leqq f_2 \\ 0 \quad \text{elsewhere} & , \end{cases}$$

and the relations between statistical errors and the amplitudes of the common noise signal may be stated as:

$$\delta R = \frac{{}_{\tau}\Delta \bar{\bar{C}}_m}{\bar{\bar{C}}_m(0)} = \frac{1}{\sqrt{(f_2 - f_1) T}} \quad ,$$

and

$$\frac{\Delta \bar{\bar{C}}_m(\tau, T)}{\bar{\bar{C}}_m(0)} = \delta R \sqrt{1 + \frac{\sin 4\pi f_2 \tau - \sin 4\pi f_1 \tau}{4\pi (f_2 - f_1) \tau}} \quad . \tag{44}$$

The first relation explains why the noise bandwidth B, which is obtained from the slope of the error curve, makes the filter bandwidth (f_2-f_1). The second relation has been used to check the variation of statistical error as a function of time lag. This comparison is discussed in detail in the section, Demonstration of the Piecewise Method.

DERIVATIVES OF THE COVARIANCE FUNCTION

The piecewise program also includes an option for the calculation of the first and second derivatives of the covariance function. In some situations these derivatives may be the physical quantity of interest in their own right. We shall also demonstrate here that they are related to the covariance of the time derivatives of the signals. Thus statistics based on the signal time derivatives rather than on the signals themselves are automatically included within this option. Finally, as we shall discuss in the next section, a knowledge of these derivatives may be employed to interpolate the covariance curves between measured lag points.

Let us consider the covariance between the delayed signal $y(t+\tau)$ and the time derivative of the $x(t)$ signal. The estimate of this covariance for the i^{th} piece is, from our previous discussion,

$$\bar{C}_i^{(1)}(\tau) = \frac{1}{\Delta T} \int_{(i-1)\Delta T}^{i\Delta T} \frac{\partial x(t)}{\partial t} [y(t+\tau) - \bar{y}_i] \, dt \quad . \tag{45}$$

The superscript 1 indicates that one factor in the instantaneous product is a time derivative. If both factors are time derivatives we obtain

$$C_i^{(2)}(\tau) = \frac{1}{\Delta T} \int_{(i-1)\Delta T}^{i\Delta T} \frac{\partial x(t)}{\partial t} \frac{\partial y(t+\tau)}{\partial t} \, dt \quad .$$

Direct calculation of $\bar{C}_i^{(1)}(\tau)$ and $\bar{C}_i^{(2)}(\tau)$ is not possible since instantaneous digital samples of the signals, not their time derivatives, are stored in the computer. We approximate the required time derivatives with a finite difference quotient between two digital samples which are separated by the sampling interval (ϵ) of the analog-to-digital converter.

$$\frac{\partial x(t)}{\partial t} \cong \frac{x(t+\epsilon) - x(t)}{\epsilon} \quad . \tag{46}$$

$$\frac{\partial y(t+\tau)}{\partial t} \cong \frac{y(t+\tau+\epsilon) - y(t+\tau)}{\epsilon} \quad . \tag{47}$$

Substituting these estimates into equation (45) and using the normal accumulative procedures over m pieces, we obtain the quantity

$$\bar{\bar{C}}_m^{(1)}(\tau) = \frac{1}{\epsilon} \sum_{i=1}^{m} \frac{1}{\Delta T} \int_{(i-1)\Delta T}^{i\Delta T} [x(t+\epsilon)\, y(t+\tau) - x(t)\, y(t+\tau)] \, dt \quad . \tag{48}$$

If now it is assumed that the data are stationary over the sampling period ϵ, then

$$\int x(t + \epsilon)\ y(t + \tau)\,dt = \int x(t)\ y(t + \tau - \epsilon)\ dt \quad .$$

Substituting this identity into equation (48) we find that

$$\bar{\bar{C}}_m^{(1)}(\tau) = \frac{1}{m} \sum_{i=1}^{m} \frac{\bar{C}_i(\tau - \epsilon) - \bar{C}_i(\tau)}{\epsilon} = -\frac{\partial}{\partial \tau}\ \bar{\bar{C}}_m(\tau) \quad . \tag{49}$$

A similar derivation yields the result for $\bar{\bar{C}}_m^{(2)}(\tau)$, namely,

$$\bar{\bar{C}}_m^{(2)}(\tau) = \frac{\partial^2}{\partial \tau^2}\ \bar{\bar{C}}_m(\tau) \quad . \tag{50}$$

We see, therefore, that the approximation of signal time derivatives by sample difference quotients equations (46) or (47) can be employed to yield:

1. The covariance between one signal and the time derivative of a second signal. This quantity is also the first derivative of the covariance function of the signals themselves.

2. The covariance between time derivatives of two signals. This quantity is identical to the second time derivative of the signal covariance function.

Finally, it should be mentioned that accumulative error analyses for the quantities $\bar{\bar{C}}_m^{(1)}(\tau)$ and $\bar{\bar{C}}_m^{(2)}(\tau)$ are also included. They follow precisely the methods used for the covariance functions as discussed previously.

Results obtained for the first and second derivatives of the covariance function of a band-limited white noise signal together with the associated error estimates are presented in the Conclusions section.

INTERPOLATION BETWEEN LAG POINTS

Although the covariances of time derivatives $\bar{\bar{C}}_m^{(1)}(\tau)$ and $\bar{\bar{C}}_m^{(2)}(\tau)$ were introduced principally for reasons of physical interpretation, their identity with the time-lag derivatives, equations (49) and (50), permits the detail of the covariance curves, $\bar{\bar{C}}_m(\tau)$, to be improved between actually calculated lag points.

Such improvement of detail is important since it can:

1. Separate the requirements for lag points and integration points which are required for calculation of Fourier transforms. Points for the Fourier integral can be provided by interpolation instead of direct calculation of lagged product mean values.

2. Minimize the number of calculated time lags by matching distributions of interpolated values to the type of detail required of the covariance curve. Any special distribution of interpolated values such as inverse $(1/\tau)$ spacing of logarithmic spacing could be prescribed.

3. Save computer time by stopping calculation as soon as statistical errors are reduced below the level of fixed numerical errors. This is discussed at the end of this section.

The above advantages may be realized by any interpolation routine which uses time-lag derivatives. We have chosen a polynomial interpolation of fifth order around the center point $\tau_{\ell + 1/2}$:

$$\bar{\bar{C}}^{(0)}_{\text{polyn}}(\tau) = \sum_{n=0}^{5} a_n \left(\frac{\tau - \tau_{\ell + 1/2}}{\tau_{\ell+1} - \tau_\ell} \right)^n \quad . \tag{51}$$

This interpolation is valid inside an interval,

$$\tau_\ell \leqq \tau \leqq \tau_\ell + 1 \quad ,$$

which is terminated by two lag points for which the covariance function $\bar{\bar{C}}_m(\tau)$ and its first and second derivative, $\bar{\bar{C}}_m^{(1)}(\tau)$ and $\bar{\bar{C}}_m^{(2)}(\tau)$, are accessible from lagged product calculations. The center of this interval is

$$\tau_{\ell+1/2} = \frac{1}{2}(\tau_{\ell+1} + \tau_\ell) \quad . \tag{52}$$

The coefficients a_n of the polynomial approximation are found by comparing the derivatives which the above interpolation would give at the end points τ_ℓ and $\tau_{\ell+1}$ with the derivatives that have actually been calculated at these points. This comparison leads to a system of six linear algebraic equations which have been grouped into two blocks of three equations. The first group of equations is

$$\sum_{n=0}^{5} \frac{n!}{(n-j)!} \left[(1/2)^n + (-1/2)^n \right] = \alpha_j \quad , \tag{53}$$

where $j = 0,\ 1,\ 2$. Here α_j describes the sum of the normalized derivatives that are given for the end points,

$$\alpha_j = \frac{\left(\tau_{\ell+1} - \tau_\ell\right)^j}{2} \left[C_m^{(j)}(\tau_{\ell+1}) + \bar{\bar{C}}_m^{(j)}(\tau_\ell) \right] \quad , \tag{54}$$

where $j = 0,\ 1,\ 2$. The normalized derivatives in α_0 and α_2 follow directly from the finite difference approximations given in equations (46) and (47). In the case of α_1, one must correct for the dislocation of the estimated $R_m^{(1)}$ value.

$$\begin{aligned} \alpha_1 = \frac{\tau_{\ell+1} - \tau_\ell}{2} \quad & \bar{\bar{C}}_m^{(1)}(\tau_{\ell+1} + \epsilon/2) - \bar{\bar{C}}_m^{(1)}(\tau_\ell + \epsilon/2) \\ & - \frac{\epsilon}{2} \left[R_m^{(2)}(\tau_{\ell+1}) - \bar{\bar{R}}_m^{(2)}(\tau_\ell) \right] \quad . \end{aligned} \tag{55}$$

The second group of equations is

$$\sum_{n=0}^{5} \frac{n!}{(n-j)!} \left[(1/2)^n - (-1/2)^n \right] = \beta_j \quad . \tag{56}$$

Here β_j describes the difference of the normalized derivatives which are given at the end points.

$$\beta_j = \frac{\left(\tau_{\ell+1} - \tau_\ell\right)^j}{2} \left[\overline{\overline{C}}_m^{(j)} (\tau_{\ell+1}) - \overline{\overline{C}}_m^{(j)} (\tau_\ell) \right] \quad . \tag{57}$$

The normalized derivatives in β_0 and β_j follow directly from the finite difference approximations given in equations (46) and (47). In using finite differences for the approximation of α_1, one must correct once again for the dislocation of the finite difference.

$$\begin{aligned} \beta_1 = {} & \frac{\tau_{\ell+1} - \tau_\ell}{2} \left[\overline{\overline{R}}_m^{(1)} (\tau_{\ell+1} + \epsilon/2) - \overline{\overline{R}}_m^{(1)} (\tau_\ell + \epsilon/2) \right] \\ & - \frac{\epsilon}{2} \left[\overline{\overline{R}}_m^{(2)} (\tau_{\ell+1}) - \overline{\overline{R}}_m^{(2)} (\tau_\ell) \right] \quad . \end{aligned} \tag{58}$$

The six values of α_j and β_j are thus known from the finite difference approximation of the time derivatives. The two sets of equations therefore represent two systems of three linear algebraic equations for the odd and even polynomial coefficients, respectively. Solving these equations for the desired polynomial coefficients gives the result:

$$\begin{aligned} a_0 &= \alpha_0 + \frac{1}{16} \left(\frac{\alpha_2}{2} - 5 \alpha_1 \right) \quad , \\ a_1 &= \frac{1}{8} \left(15\beta_0 - \frac{7}{2} \beta_1 + \frac{1}{4} \beta_2 \right) \quad , \\ a_2 &= \frac{1}{2} \left(3\alpha_1 - \frac{1}{2} \alpha_3 \right) \quad , \\ a_3 &= 10 \left(\frac{1}{2} \beta_1 - \beta_0 - \frac{1}{20} \beta_2 \right) \quad , \\ a_4 &= \frac{1}{2} \alpha_2 - \alpha_1 \quad , \\ a_5 &= 12 \left(\beta_0 - \frac{1}{2} \beta_1 - \frac{1}{12} \beta_2 \right) \quad . \end{aligned} \tag{59}$$

Substituting these coefficients into the polynomial of equation (51), we have thus derived the interpolation of the covariance $\bar{\bar{C}}_m^{(0)}(\tau)$ function between two lag points, $\tau_{\ell+1}$ and τ_ℓ. This interpolation is graphically illustrated in Figure 2. The known values of $\bar{\bar{C}}_m^{(0)}$, $\bar{\bar{C}}_m^{(1)}$, and $\bar{\bar{C}}_m^{(2)}$ at the two end points are indicated by the ordinate, the slope, and the radius of curvature at these points. The polynomial interpolation will then match these ordinates, slopes, and curvatures as indicated by the dashed curve. The interpolation formula of equation (51) can be used to calculate any number of points of this dashed curve. We have largely used 5 to 10 points.

An interpolation routine is useful only if one knows some limits for the error between the expected curve $\bar{\bar{C}}_m^{(0)}$ and its interpolated value $C_{\text{polyn.}}^{(0)}$ In the preceding interpolation there are several types of errors. First, there are truncation errors which account for the fact that the polynomial of equation (51) is a Taylor series approximation around the midpoint, which has been truncated at the fifth term. However, additional errors enter the interpolation since the calculated values of α_j and β_j are inflicted by statistical errors $\Delta\bar{\bar{C}}_m^{(j)}(\tau_\ell)$ and $\Delta\bar{\bar{C}}_m^{(j)}(\tau_{\ell+1})$ as well as quantization errors of the finite difference approximations. We believe, therefore, that a detailed error analysis is hopelessly complicated.

The desire to judge the validity of the polynomial approximation has led to a simpler approach, which uses an interpolation amplitude, Δ_ℓ. This amplitude accounts for the difference between the above polynomial interpolation and a straight-line interpolation which does not depend on slopes and curvatures.

$$\bar{\bar{C}}_m^{(0)}(\tau)_{\text{straight}} = 2\,\frac{\tau - \tau_{\ell+1/2}}{\tau_{\ell+1} - \tau_\ell}\,\beta_0 + \alpha_0 \quad . \tag{60}$$

The interpolation amplitude is now defined as the mean square difference between the polynomial and the straight-line interpolation.

$$\Delta_\ell^2 = \frac{1}{\tau_{\ell+1} - \tau_\ell} \int_{\tau_\ell}^{\tau_{\ell+1}} \left[\overline{\overline{R}}^{(0)}_{\text{polyn}}(\tau) - \overline{\overline{R}}^{(0)}_{\text{straight}}(\tau) \right]^2 d\tau$$

$$= \frac{1}{\tau_{\ell+1} - \tau_\ell} \int_{\tau_\ell}^{\tau_{\ell+1}} \left[\sum_n a_n \left(\frac{\tau - \tau_{\ell+1/2}}{\tau_{\ell+1} - \tau_\ell} \right)^n - 2 \frac{\tau - \tau_{\ell+1/2}}{\tau_{\ell+1} - \tau_\ell} \beta_0 - \alpha_0 \right]^2 d\tau$$

$$= \frac{1}{\tau_{\ell+1} - \tau_\ell} \int_{\tau_\ell}^{\tau_{\ell+1}} \left[\sum_{n=1}^{5} \overline{a}_n \left(\frac{\tau - \tau_{\ell+1/2}}{\tau_{\ell+1} - \tau_\ell} \right)^n \right]^2 d\tau \qquad (61)$$

$$= \sum_{n=0}^{5} \frac{\overline{a}_n^{\,2}}{(2n+1)\,4^n} + \frac{1}{6} \overline{a}_0 \overline{a}_2 + \frac{1}{40} (\overline{a}_0 \overline{a}_4 + \overline{a}_1 \overline{a}_3)$$

$$+ \frac{1}{112} (\overline{a}_1 \overline{a}_5 + \overline{a}_2 + \overline{a}_4) + \frac{1}{558} \overline{a}_3 \overline{a}_5 \quad .$$

The coefficients $\overline{a}_n$ are related to the polynomial coefficients by:

$$\overline{a}_0 = a_0 - \alpha_0$$

$$\overline{a}_1 = a_1 - 2\beta_0 \qquad (62)$$

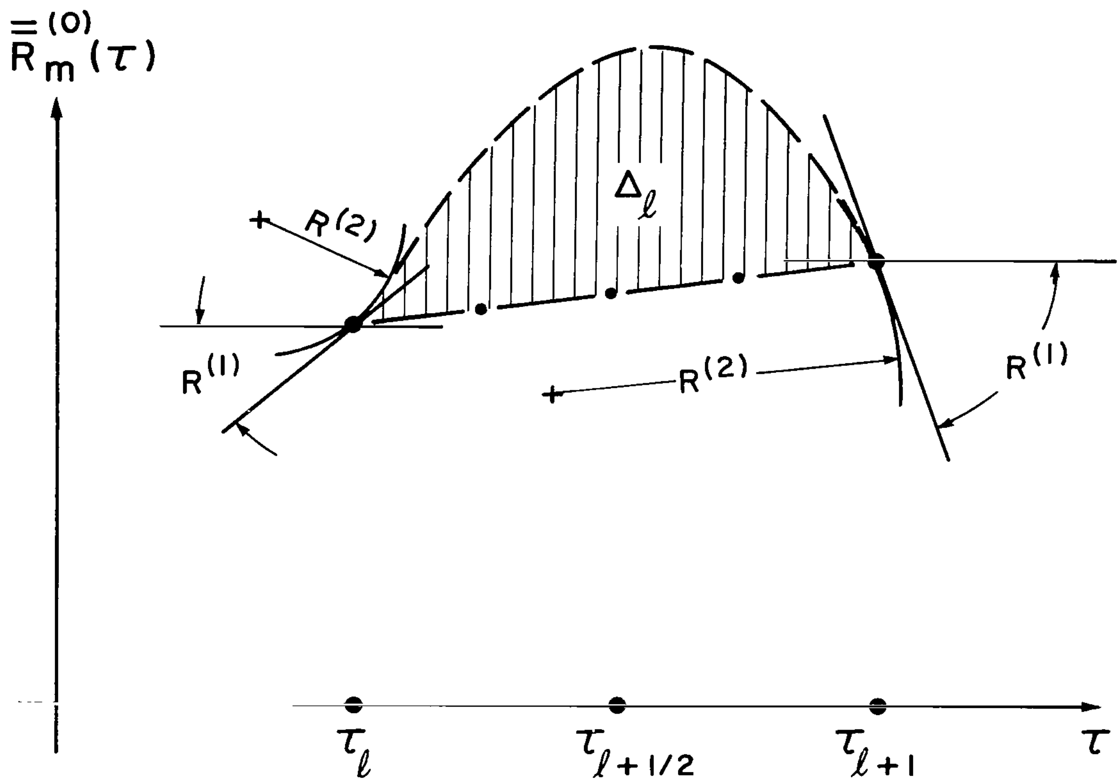
$$\overline{a}_n = a_n \text{ for } n \geqq 2 \quad ,$$

and can thus be calculated directly.

The interpolation amplitude is a measure for the effect which the use of first and second derivatives has on improving a straight-line interpolation. The polynomial approximation will obviously become meaningless if the interpolation amplitude becomes smaller than the statistical error $\Delta \overline{\overline{C}}_m^{(0)}$ at the end points, τ_ℓ and $\tau_{\ell+1}$. It will also be meaningless if the end points,

— — — Polynomial $\sum a_n \left(\frac{\tau - \tau_{\ell+1/2}}{\tau_{\ell+1} - \tau_{\ell}}\right)^n$

—•—•— Straight $2\,\frac{\tau - \tau_{\ell+1/2}}{\tau_{\ell+1} - \tau_{\ell}}\,\beta_0 + \alpha_0$



$$\Delta_\ell^2 = \frac{1}{\tau_{\ell+1} - \tau_\ell} \int_{\tau_\ell}^{\tau_{\ell+1}} \left(R^{(0)}_{Polyn.} - R^{(0)}_{Straight}\right)^2 d\tau$$

Figure 2. Graphical illustration of interpolation routine.

τ_ℓ and $\tau_{\ell+1}$, are spaced too far apart. For example, if the estimated covariance $\bar{\bar{C}}_m^{(0)}(\tau)$ has more than one extreme between τ_ℓ and $\tau_{\ell+1}$, the second derivative will oscillate significantly, and the polynomial interpolation is subject to large truncation errors. Such errors may be anticipated when the interpolation amplitude is equal to or larger than the peak value of $\bar{\bar{R}}_m^{(0)}(\tau)$, which is indicated by the lagged product calculations in the vicinity of the interpolation interval. The validity of the polynomial approximation can thus be judged to some extent by comparing the interpolation amplitudes with the statistical error $\Delta \bar{\bar{C}}_m^{(0)}$ and/or the indicated peak values of the directly calculated covariance estimates $\bar{\bar{C}}_m^{(0)}(\tau)$.

For any given set of data the errors caused by the truncation of the Taylor series and by the finite difference approximation of derivatives are fixed as soon as the sampling period ϵ and the distribution of lag points τ_ℓ have been chosen. It does not make sense to reduce the statistical error below these fixed numerical errors. In view of these considerations, we have introduced a third option to stop the piecewise operations. This option is based on the average interpolation amplitude, $\bar{\Delta}$, which is taken as some indirect measure for the combined numerical errors. The word "average" refers to an integral over the time lag range, $-\tau_{N+1} \leqq \tau \leqq \tau_{N+1}$. A stepwise approximation of this integral is

$$\Delta = \left[\frac{1}{2\tau_{N+1}} \sum_{\ell=-N+1}^{n} (\tau_{\ell+1} - \tau_\ell)\, \Delta_\ell^2 \right]^{1/2} . \tag{63}$$

The program will stop accepting new data pieces if the average statistical error has been reduced below the average interpolation amplitude, $\bar{\Delta}$. However, this option will not be used if a statistical error other than zero is specified by the user.

A demonstration of this interpolation routine is given in the next section, Demonstration of the Piecewise Method.

DEMONSTRATION OF THE PIECEWISE METHOD

The objective of this section is to demonstrate the operation and special features of the piecewise program, which were reviewed in the earlier sections. To permit meaningful discussion of results, a relatively simple test signal, namely, band-limited white noise, was used. Special precautions (outlined below) insured to the largest possible extent, that uncertainties in the results were not introduced by the nonideal nature of the test signal itself.

In this section we shall demonstrate from the results the existence of the subdivision error and its relative ineffectiveness as a high-pass filter. The dependence of the statistical uncertainty of the results on both bandwidth and integration time is also demonstrated. A discussion is presented of the relative merits of using the interpolation routine, that is, calculating relatively few points on the correlation curve and using the first and second derivatives at these points to complete the curve, as opposed to the direct calculation of the detailed curve. Finally, to demonstrate the potential accuracy of the complete system, a Fourier transformation of the cross-correlation curves obtained is undertaken, and the resulting spectrum is compared with the transfer function of the filter employed in generating the test signal.

The Test Signal

The test signal used for these tests was band-limited white noise, which was generated by passing the output of a white noise generator through an audio frequency spectrometer. Specifically, the signal was passed through the one-octave band filter centered at 1 kHz. According to the manufacturer's manual, the -3 dB points on this filter are at 700 and 1400 Hz, respectively. This form of signal was chosen because:

1. It provides a test of the system in terms of a random, rather than deterministic, (i.e., sine wave) input signal.

2. The use of a "fixed" filter offers a greater reliability of spectrum shape then can generally be obtained from variable band-pass filters.

3. The response function of the filter is available in the manufacturer's operation manual and can be used as an independent check of the results.

4. The parameters of the signal, notably its bandwidth, are unambiguously defined.

5. A center frequency of 1 kHz, together with a conveniently available analog-to-digital conversion rate of 20 000 samples per second, offers a reasonable degree of resolution of the correlation curve as a function of time delay.

The test method used was to generate a single signal as outlined here and to pass this to both channels of the analog-to-digital converter, thus generating two identical signals on the resulting digital tape. Cross-correlation calculations were then made on these two signals, resulting in the generation of the autocorrelation of the original band-limited signal. Using this method does provide a test of the analog-to-digital converter and computer program but keeps to the irreducible minimum any dependence of the results on the other peripheral equipment used to generate that signal.

A typical example of the correlation function obtained is shown in Figure 3 for future reference purposes.

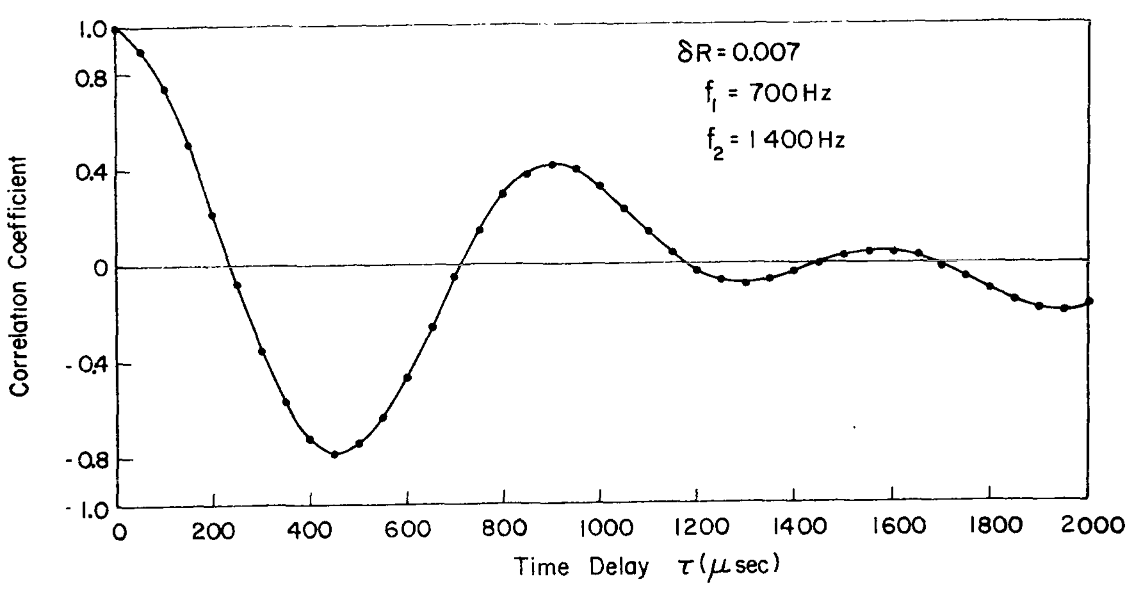


Figure 3. Correlation curve for band-limited white noise.

Subdivision Error

An analysis demonstrating the possible occurrence of subdivision error when using the piecewise method for calculating correlation functions was presented in the section, Accumulation of Piecewise Averages. As shown there, the effect arises because the amplitude of fluctuations is calculated relative to the mean value of the piece in which they are contained, not relative to the mean value of the entire record. This results in the partial elimination of the power contained in fluctuations whose period is in excess of the piece length.

Let us next consider the effect of subdivision error on the results of our present tests with the band-limited white noise signal. Let us suppose initially that the use of a piece length ΔT eliminates <u>all</u> energy contained in the test signal below a frequency f_c where f_c and ΔT are related by the expression

$$f_c = \frac{r}{\Delta T} \quad , \tag{64}$$

where r is a number, probably of order unity. Furthermore, the cross-correlation function of two signals is related to their cross-power spectral density function by the Fourier transform relationship:

$$\bar{\bar{C}}_m(\tau) = \int_0^\infty S_{xy}(\omega)\, e^{-i\omega\tau}\, d\omega \quad . \tag{65}$$

However, since for the test example used here

$$x(t) = y(t) \quad ,$$

equation (2) can be written

$$\bar{\bar{C}}_m(\tau) = \int_0^\infty S_{xx}(\omega)\, e^{-i\omega t}\, d\omega \quad , \tag{66}$$

while for the particular case $\tau = 0$,

$$\bar{\bar{C}}_m(0) = \int_0^\infty S_{xx}(\omega)\, d\omega \quad . \tag{67}$$

Equation (67) represents the well-known fact that the correlation at zero time delay is equal to the total signal energy.

If, however, the use of a piece length T were to eliminate all energy below a frequency f_c, equation (65) would become

$$C_m(0, \Delta T) = \int_{2\pi f_c}^\infty S_{xx}(\omega)\, d\omega \quad . \tag{68}$$

Thus if the spectral density function of the signal $S_{xx}(\omega)$ is known, it is possible to calculate $R_{xx}(0, \Delta T)$ as a function of the piece length ΔT.

For the present test signal a useful approximation of the spectral density function is

$$S_{xx}(\omega) = 1 \quad 2\pi f_1 \leqq \omega \leqq 2\pi f_2$$

$$= 0 \quad \text{elsewhere} \quad ,$$

where

$$f_1 = 700 \text{ Hz} \quad \text{and} \quad f_2 = 1400 \text{ Hz} \quad .$$

The decrease of $\bar{\bar{C}}_m(0, \Delta T)$ with decreasing piece length calculated on the basis that all energy below f_c is eliminated for two values of r [equation (64)] is shown in Figure 4. Also shown for comparison are the values of $\bar{\bar{C}}_m(0, \Delta T)$ obtained from the piecewise correlation program. These data were generated by computing the correlation curve for various piece lengths in the range $450 \leqq \Delta T \leqq 12\ 000$ μsec. In all cases an integration time of 30 sec was employed, resulting in a statistical uncertainty of the order of 0.7 percent.

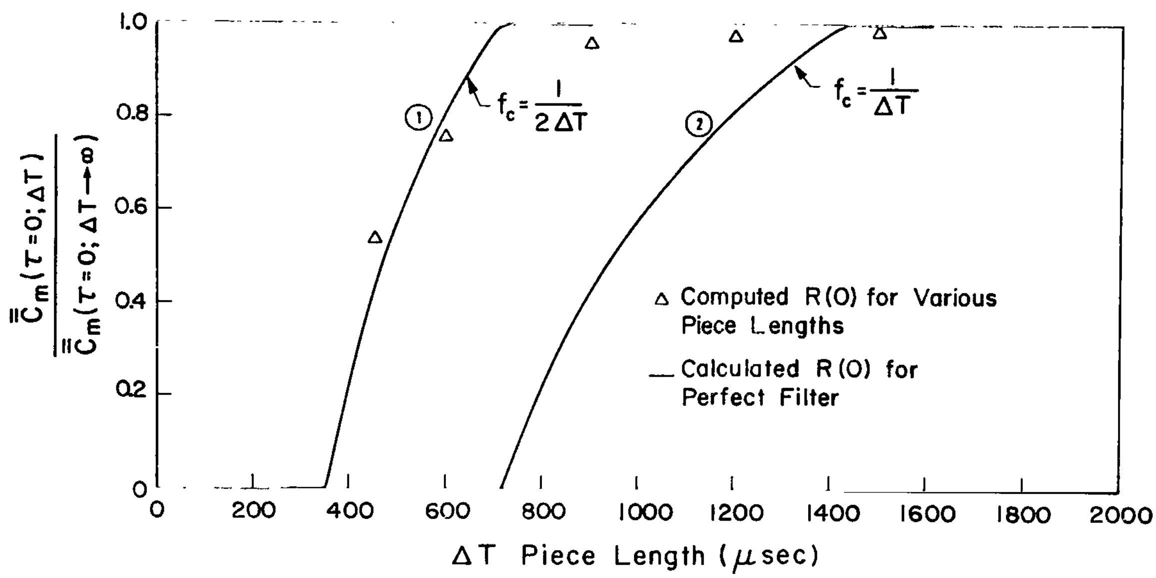


Figure 4. Effect of piece length on apparent signal power.

Although the experimental results clearly indicate the expected decrease of the apparent signal power, $\overline{\overline{C}}_m(0, \Delta T)$, with decreasing piece length, the rate of decrease is appreciably less than that obtained theoretically where it was assumed that all energy below the frequency f_c was eliminated. However, partial elimination of energy below a frequency f_c where

$$f_c = \frac{1}{\Delta T}$$

is indicated by the results. Therefore, we conclude that the use of a finite piece length will lead to an underestimate of the power contained in fluctuations whose period is in excess of the piece length.

To ascertain the relative effectiveness of piece length as a high-pass filter, the data of Figure 4 are replotted in Figure 5 as a function of the cutoff frequency f_c where the relationship,

$$f_c = \frac{1}{\Delta T} \quad ,$$

has been used.

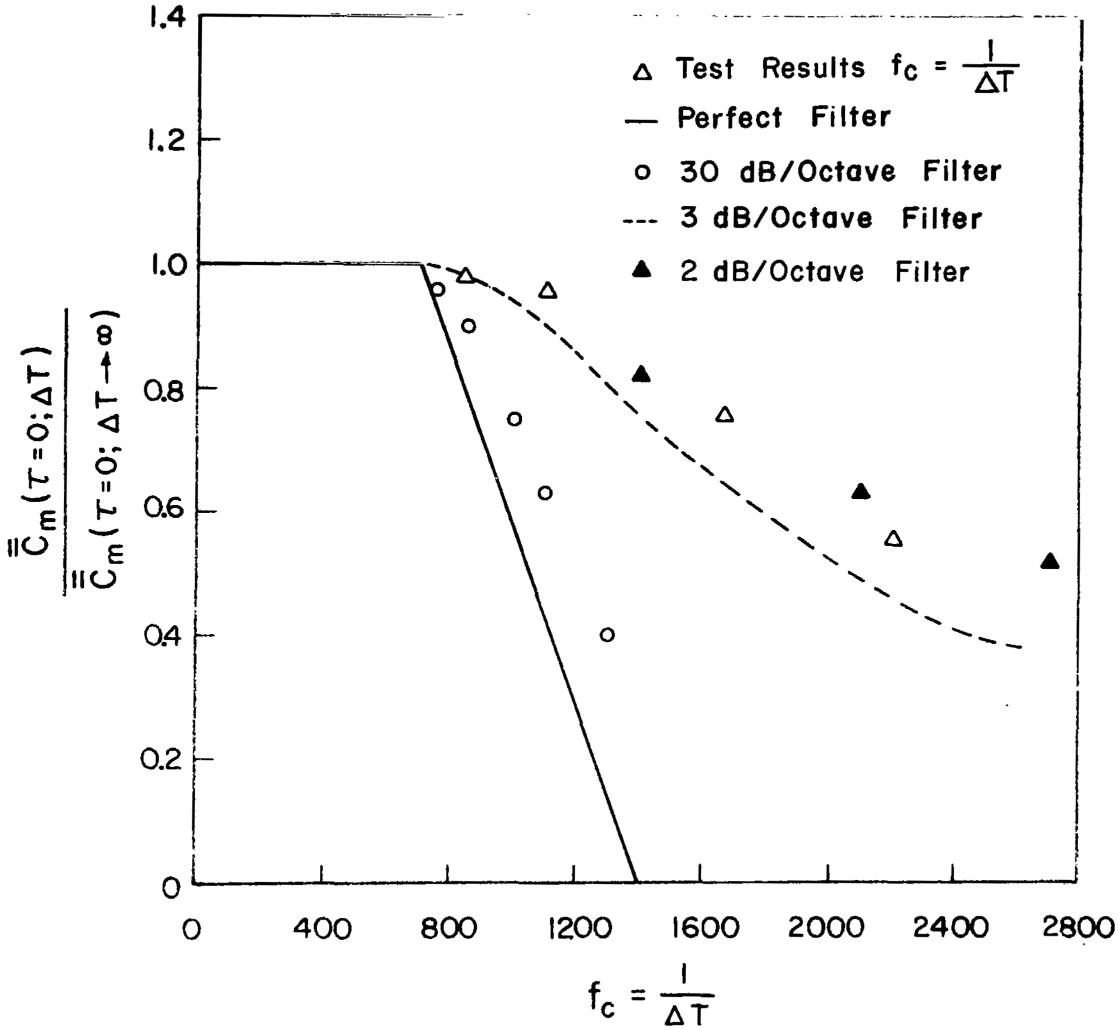


Figure 5. Effectiveness of piece length on apparent signal power.

Shown for comparison are the expected decreases of apparent energy for:

1. A perfect filter, that is, a filter which eliminates all energy below the preset frequency f_c.

2. A filter with an attenuation rate of 30 dB per octave below the preset frequency f_c.

3. A filter with an attenuation rate of 3 dB per octave.

4. A filter with an attenuation rate of 2 dB per octave.

The results indicate that the filtering capacity of the piece length is approximately comparable to that of the last-named filter. Since the attenuation rate of a good commercial filter is normally of the order of 30 dB/octave, it does not appear that filtering by choice of piece length offers a comparable alternative.

Our tests related to the subdivision error therefore lead to the following conclusions:

1. The use of finite piece lengths will lead to an underestimate of the power contained in fluctuations whose period is in excess of the piece length.

2. The resulting filtering action is rather weak, being approximately comparable to that of an analog high-pass filter with an attenuation rate of 2 to 3 dB/octave.

3. The filtering action of piece length does not offer a realistic substitute for analog high-pass filters if sharp frequency discrimination is required. It does, however, provide a convenient method for the elimination of long-term drifts, whose period is considerably in excess of those of the principal fluctuations of interest.

Statistical Accuracy

As discussed in the section, Derivatives of the Covariance Function, one of the more unique features of the piecewise program is that not only are the statistical uncertainties of the covariance estimates calculated, but also the values so derived are used as a program control parameter. Specifically, the processing of pieces continues until a specified confidence level is obtained or the data record is exhausted.

The work of Jayroe and Su [3] yields theoretical evidence regarding the behavior of these statistical errors, and we shall base our presentation of the test case results on their equation, which for the autocorrelation case may be written

$$\overline{\overline{\Delta}}_m^{\,2}(\tau, T) = \frac{1}{T} \int_{-\infty}^{\infty} S_{xx}^2(f)\, df + \frac{1}{T} \int_{-\infty}^{\infty} S_{xx}^2(f)\, e^{-i 2 \tau f}\, df \quad . \tag{69}$$

The spectrum function for the band-limited white noise may be idealized by the relation

$$S(f) = \frac{\overline{\overline{G}}_m(\tau=0)}{f_2 - f_1} \qquad f_1 \leqq f \leqq f_2$$

$$= 0 \quad \text{elsewhere} \quad .$$

Substitution of this expression into equation (69) yields

$$\frac{\overline{\overline{\Delta}}_m^{\,2}(\tau, T)}{\overline{\overline{C}}_m^{\,2}(\tau=0)} = \frac{1}{(f_2-f_1)\,T}\left[1 + \left(\frac{\sin 2\pi f_2\tau - \sin 2\pi f_1\tau}{\tau}\right)\right] \quad . \tag{70}$$

The normalized statistical error is thus seen to be composed of two items, the first independent of the time delay and a second which varies with this parameter. Furthermore, consideration of the autocorrelation function for an ideal band-limited white noise signal will show that it is of precisely the form of this second item. Thus we except the normalized error when plotted as a function of time delay to follow, in general, the trends of the autocorrelation function. We shall see shortly that this is the case, but note that this is a feature of the particular spectrum used, not something to be expected in general.

First, however, we must note that the predicted variation of statistical error with time delay requires some decision with regard to which value should be employed as the program "stop" parameter. In practice, we have chosen to average the normalized statistical error, equation (70), over the complete time-delay range computed and to utilize this average.

Specifically, the parameter δR is defined as

$$\delta R = \frac{1}{2\tau_m} \int_{-\tau_m}^{\tau_m} \frac{\overline{\overline{\Delta}}_m(\tau, T)}{\overline{\overline{C}}_m(\tau=0)}\, d\tau \quad . \tag{71}$$

The formal derivation of δR can be made using equation (70). However, the relationship is so simple that no formal derivation will be undertaken. The first term, being independent of time delay, averages to its value at any time-delay value. The second, being damped in an oscillatory manner, contributes little to the average. Thus we find that

$$\delta R = \frac{1}{\sqrt{(f_2 - f_1)T}} \quad . \tag{72}$$

To demonstrate this dependence of δR on integration time, the correlation function of the test signal described above was determined for various values of δR in the range,

$$0.15 \geqq \delta R \geqq 0.007 \quad ,$$

with a confidence factor of $p = 90$ percent. In the majority of cases the correlation function was calculated in the range $0 \leqq \tau \leqq 4000\ \mu$sec. A tabulation of the results obtained is shown in the following table.

Integration Time (sec)	Average Statistical Uncertainty (δR)
0.072	0.138
0.132	0.094
0.516	0.050
2.016	0.025
16.06	0.01
30.01	0.007

Plotting δR as a function of the square root of the reciprocal of the integration time yields the result shown in Figure 6. The results show very acceptable agreement with the prediction of equation (72). Also shown in this figure are the limits of uncertainty on the statistical error as defined by the χ^2 distribution in the section, Environmental and Statistical Variation of Piecewise Averages. The fact that all measured values of the uncertainty fall inside these limits indicates that the test signal did, indeed, exhibit the hoped-for stationarity.

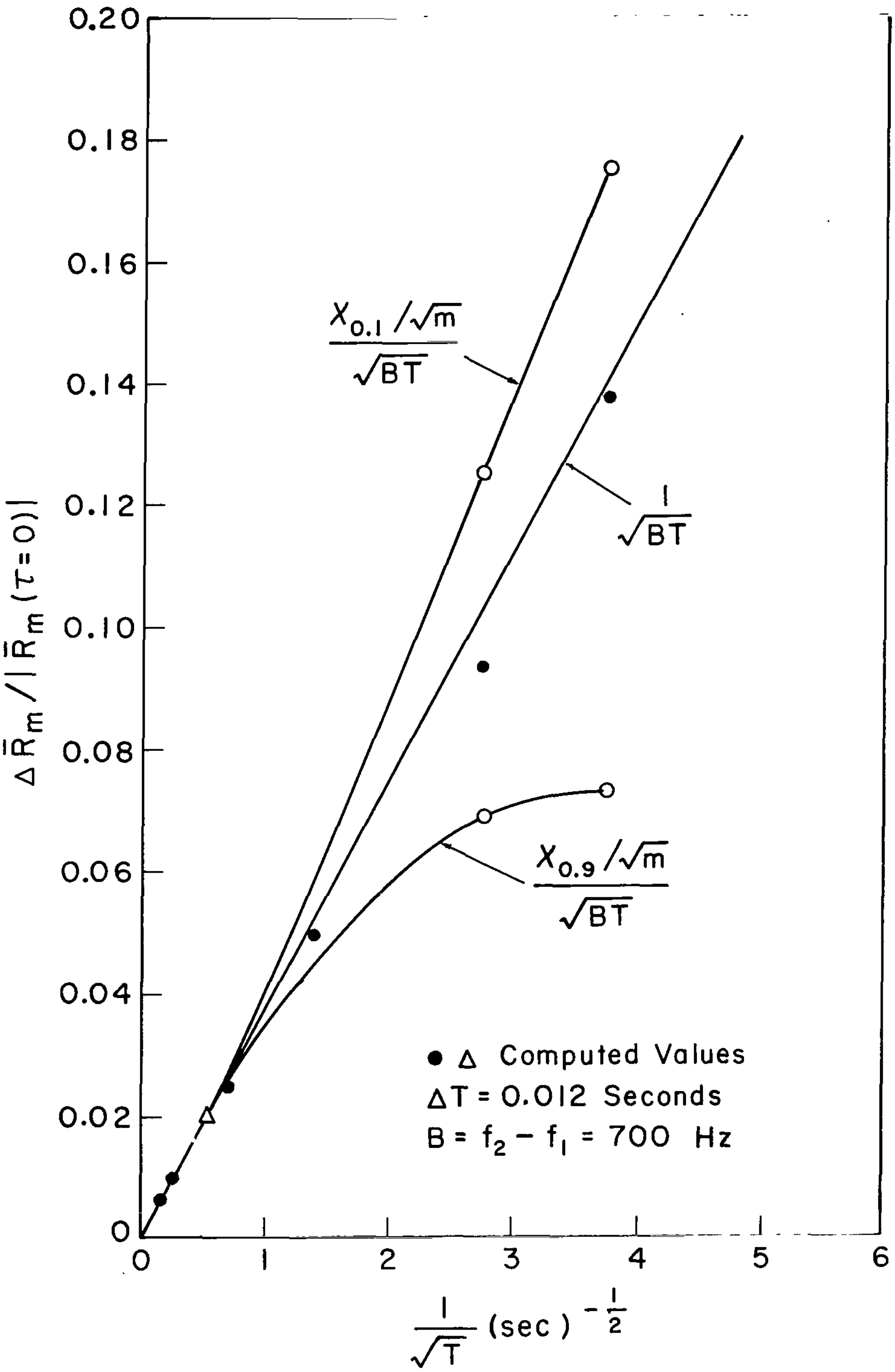


Figure 6. Accumulative statistical errors.

The results of the variation of the uncertainty with time lag are shown in Figure 7. As predicted by equation (70), the values tend to oscillate around the average value, whereas comparison with the sample autocorrelation curve of Figure 3 shows that these oscillations follow the same general trends as the autocorrelation curve as predicted above. The prediction of equation (70) is also shown superimposed on these results, and again acceptable agreement results.

Reassuring as these comparisons may be, they do not answer the question of prime importance: Do the calculated confidence intervals offer a realistic measure of the difference between the true and measured correlation curves?

To offer a precise comparison it would be necessary to have a knowledge of the exact correlation curve. Obtaining this curve is clearly not practical since its calculation would require infinite integration time. However, to obtain some estimate of the adequacy of the calculated confidence intervals, the correlation curves calculated for various integration times have been compared with the curve calculated for the maximum integration time employed, 30 sec, for which an average confidence interval of 0.007 was obtained. This latter curve is thus regarded for present purposes as the correct curve.

Specifically, the value of the cross-correlation for a particular time delay τ, obtained using an integration time T, was subtracted from that obtained for the maximum integration time T_{max}. The absolute value of this quantity was then plotted against the calculated confidence interval. The results are shown in Figure 8.

The 45-deg line drawn on this diagram corresponds to the locus of points for which

$$|\bar{\bar{C}}_m(\tau, T) - \bar{\bar{C}}_m(\tau, T_{max})| = \bar{\bar{\Delta}}_m(\tau, T) \quad .$$

Thus points plotted above and to the left of this line represent cases for which the difference between the calculated curve $\bar{\bar{C}}_m(\tau, T)$ and the correct curve is less than the calculated confidence interval. Conversely, points to the right and below the line are those for which the difference of the curves

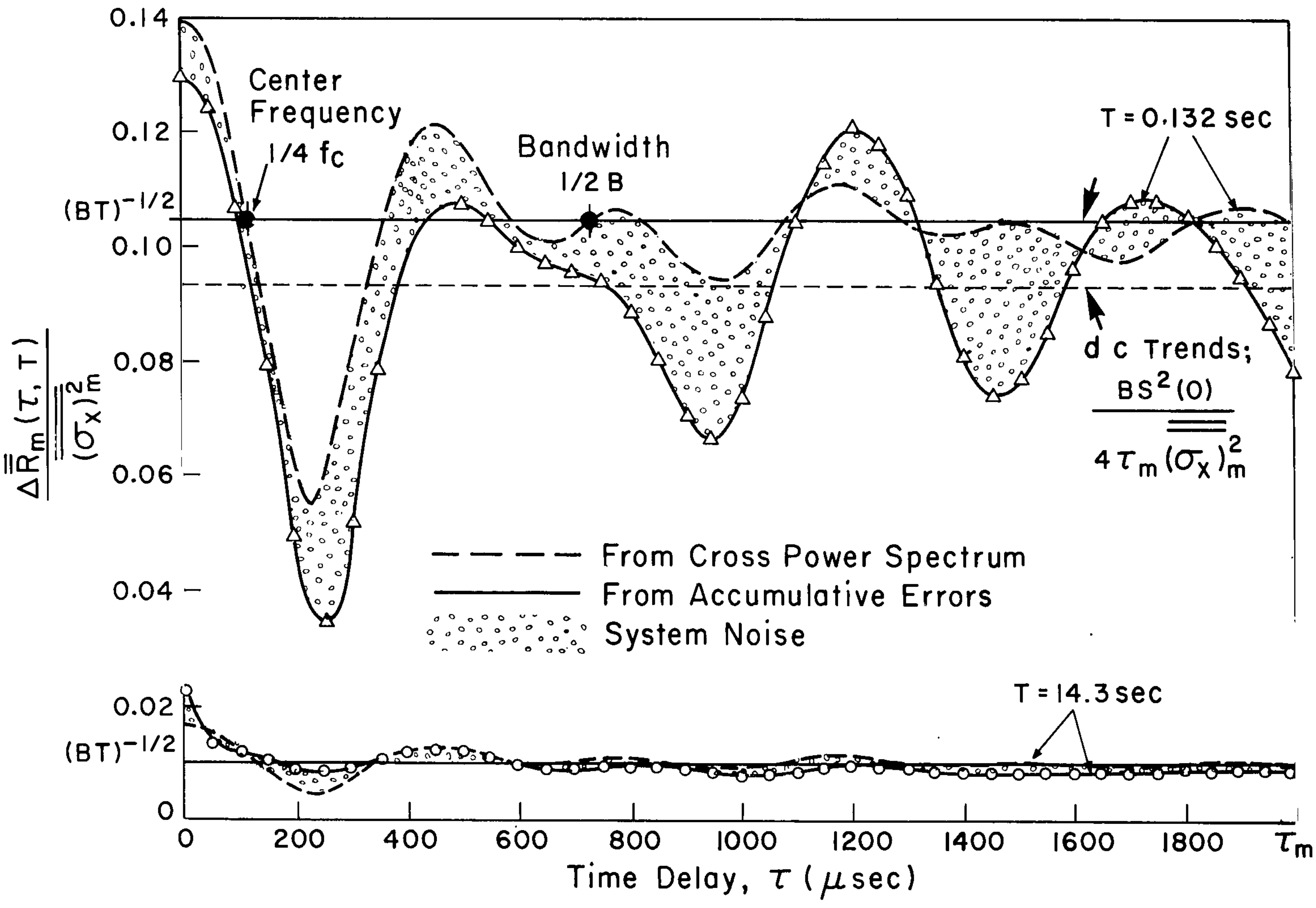


Figure 7. Statistical error of correlation curves.

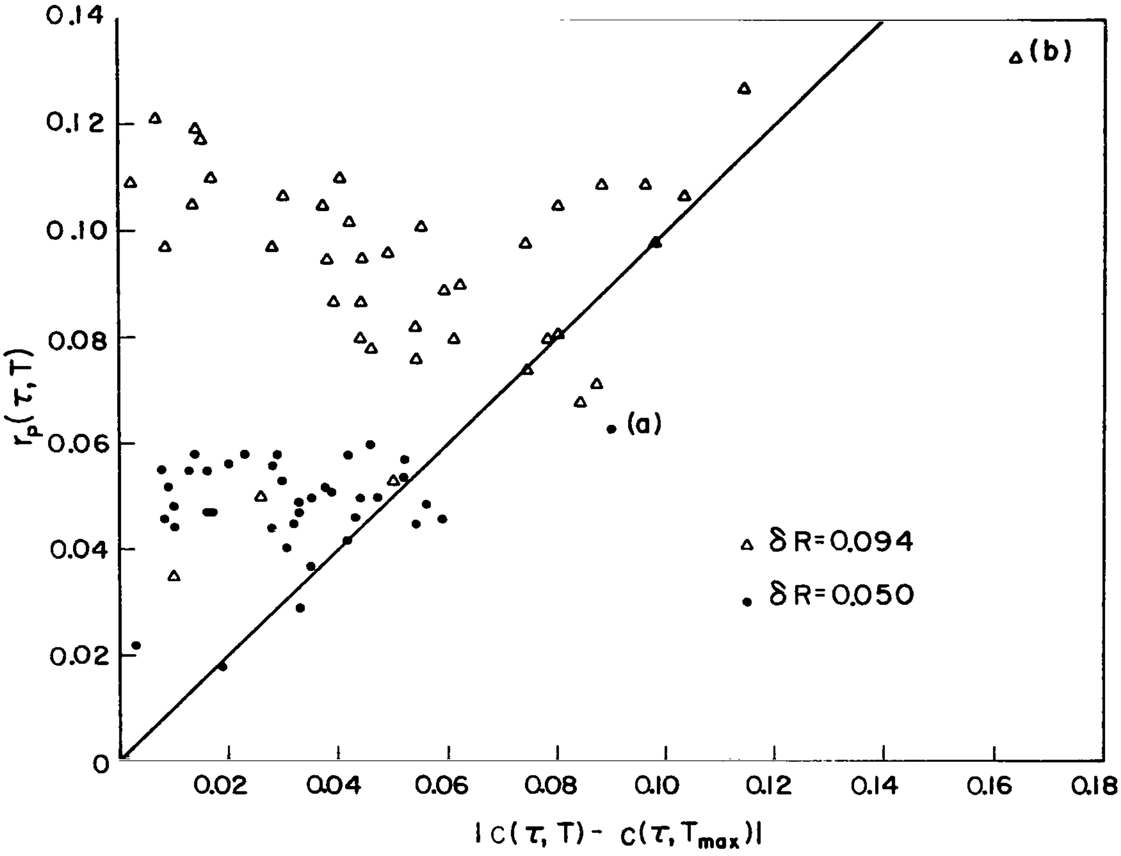


Figure 8. A comparison of estimated and actual errors of correlation measurement.

exceeds the confidence interval. In Figure 8, a total of 9 points out of a total of 80 plotted lies below the 45-deg line. Thus, for 89 percent of the points considered, the difference between the calculated and correct correlation curve is less than or equal to the confidence interval. Since this test was conducted using a probability factor of p = 90 percent, the results confirm the implied criterion that there is a 90-percent probability that the difference between the measured and true correlation is less than the confidence interval.

Covariance Derivatives and Interpolation

The calculation of the first and second derivatives of the covariance function, together with the method by which these values are used to interpolate correlation curves between measured lag points, was discussed in the sections, Derivatives of the Covariance Function and Interpolation Between Lag Points, respectively.

Figures 9 and 10 show results of the computation of the first and second derivatives, respectively, of the covariance function of one band-limited white noise test signal. Also shown for comparison are analytical calculations for these derivatives based on an ideal filter shape (Fig. 11). Although some scatter is apparent, particularly for the second derivative, the amount of agreement is encouraging. This is particularly true when one considers the alternative method of calculation, the use of finite differences between calculated covariance values.

The next question to be considered is the use of these derivatives for interpolation of covariance curves using the fifth-order polynomial approximation discussed in the section, Interpolation Between Lag Points. To make this test as meaningful as possible, the following approach was adopted. A covariance curve was first calculated using closely spaced (50 μsec) lag intervals, the calculation being continued until a certain averaged confidence interval was obtained. The calculation was then repeated, to the same confidence interval, for a set of more broadly spaced lag intervals, and the interpolation routine was used to provide detail between these latter points.

Figure 12 shows one example in which the covariance function was determined for relatively broad (500 μsec) intervals, and a further 10 points were interpolated into the intervening spaces. The averaged confidence interval for both the noninterpolated curve and the remotely spaced lag points was 0.02. The difference between the interpolated and noninterpolated points frequently exceeds this confidence interval, indicating that the numerical errors are predominant. However, the agreement obtained is acceptable for many applications, while the broad spacing of the lag points, rather less than one per turning point of the curve, should be noted.

The degree of improvement obtained by halving the distance between calculated lag points is shown in Figure 13. Indeed, deviations between the interpolated and noninterpolated curves are reduced by approximately a factor of two, but these deviations still exceed the confidence interval.

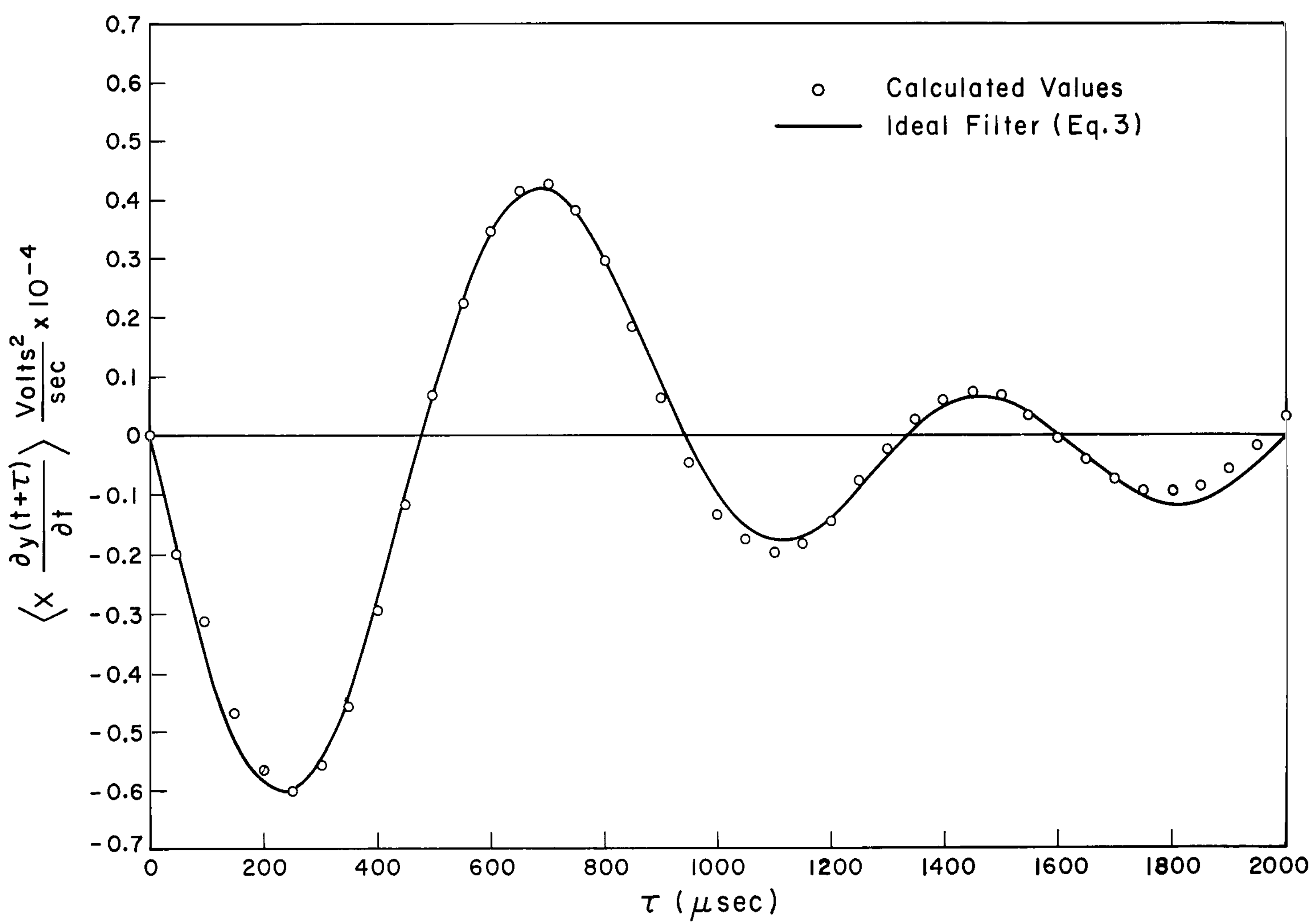


Figure 9. Correlation of signal and its time derivative.

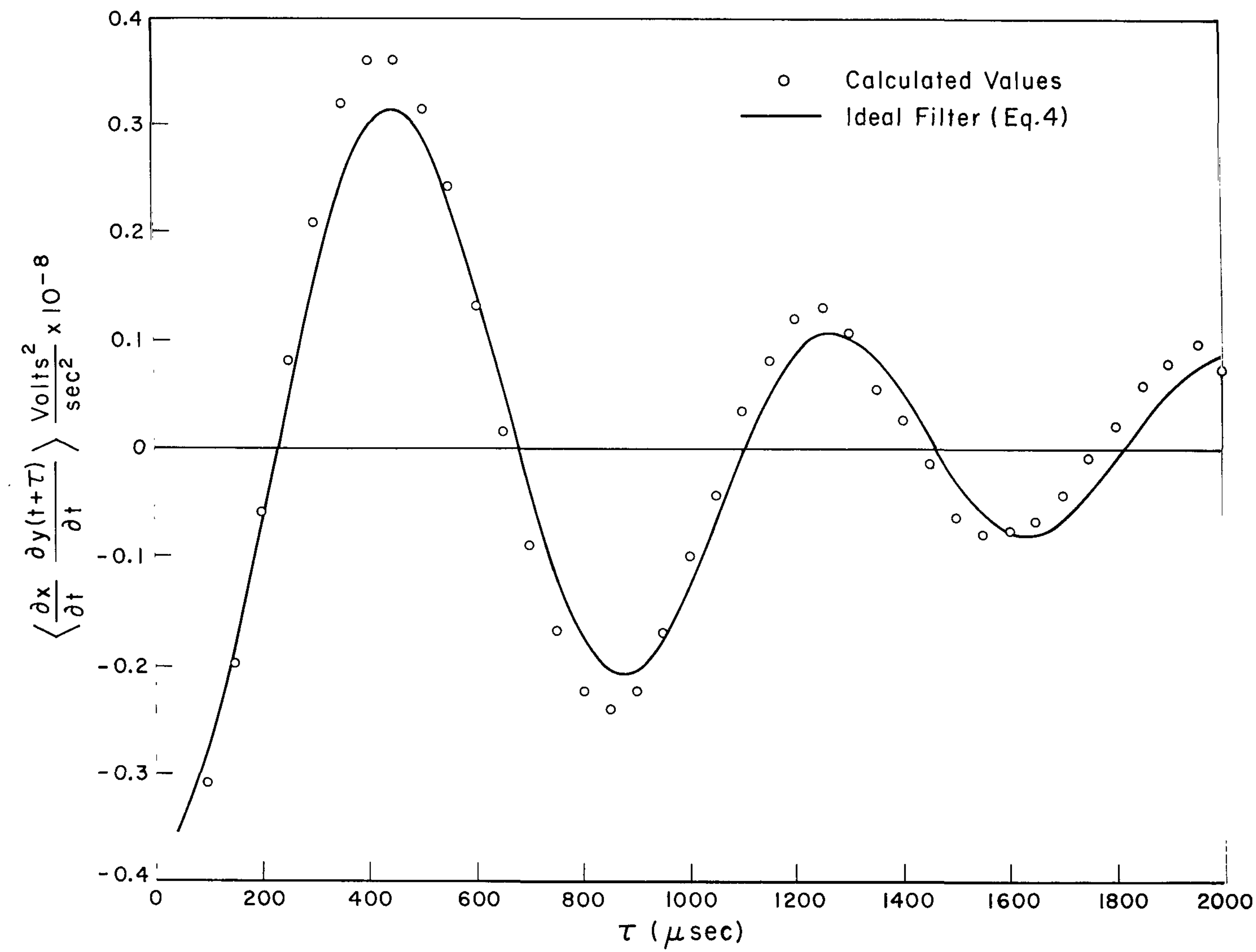


Figure 10. Correlation of time derivatives.

Finally, shown in Figure 14 is the result obtained when the averaged confidence interval was reduced to 0.007, utilizing the full 30-sec data record. Again, differences typically exceed the confidence interval.

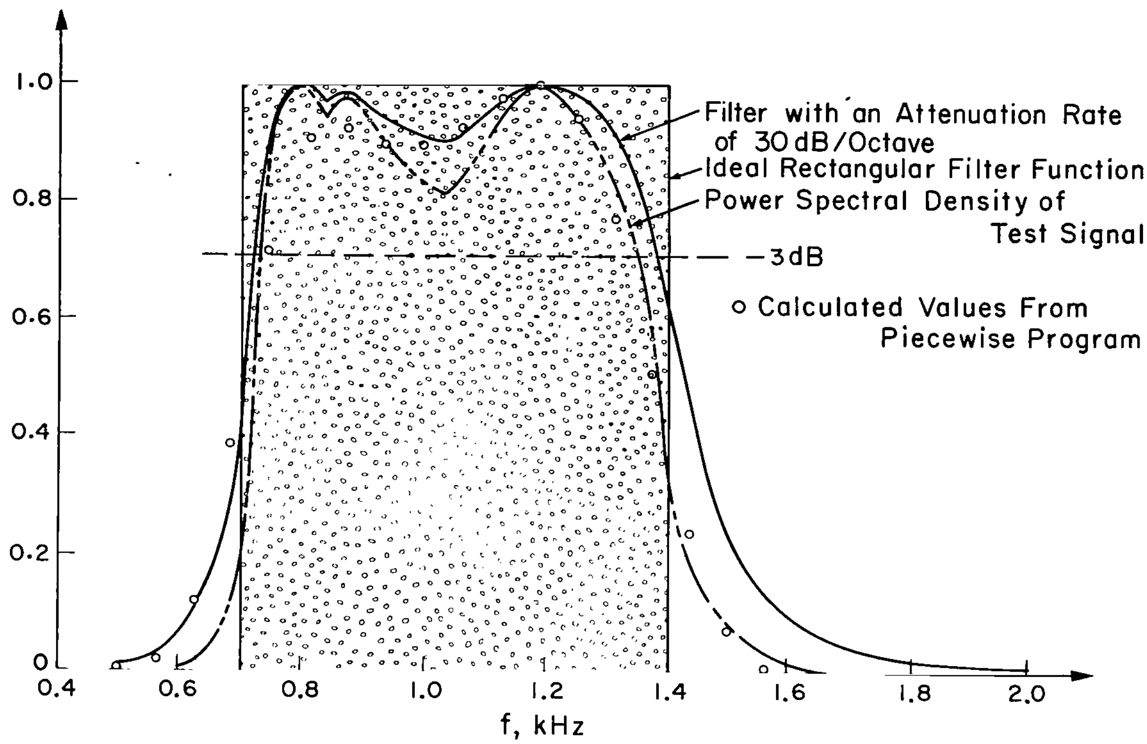


Figure 11. Normalized filter functions and power spectral density functions for ideal filters and a filter with an attenuation rate of 30 dB per octave.

On the basis of these results it is concluded that in cases where the detailed variation of a covariance curve as a function of time lag is important, direct calculation of closely spaced lag points offers the greater reliability. On the other hand, for situations where only the general variation of the curve is required, significant computer time may be saved by using interpolation routines.

The fact that the derivative calculations, available in this program, may also be used to determine the statistical properties of derivations of recorded data offers an additional valuable feature in this computer code.

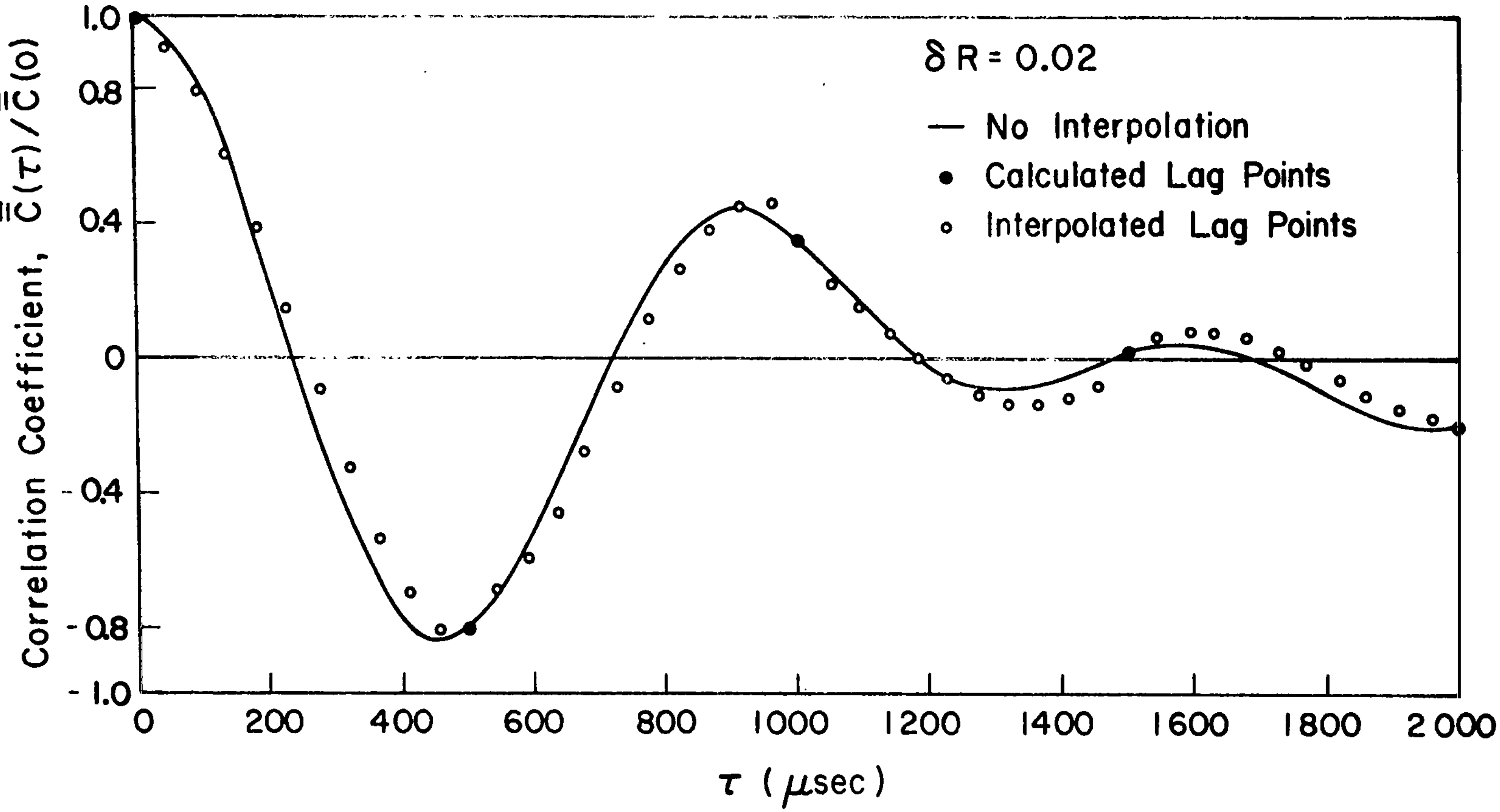


Figure 12. An interpolated correlation curve.

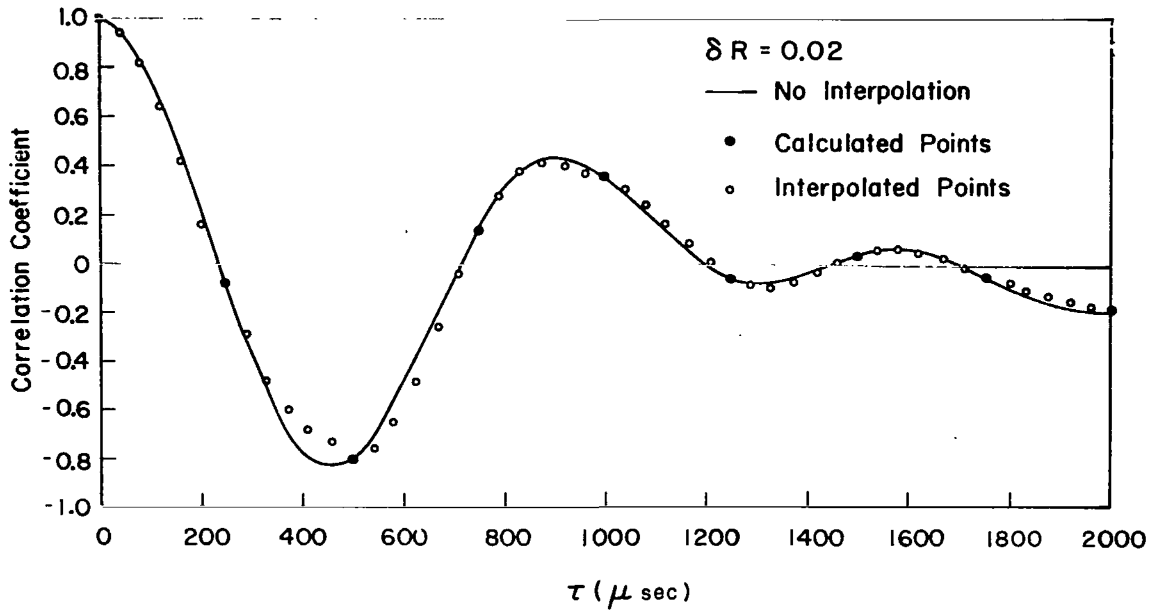


Figure 13. An interpolated correlation curve; δR = 0.02.

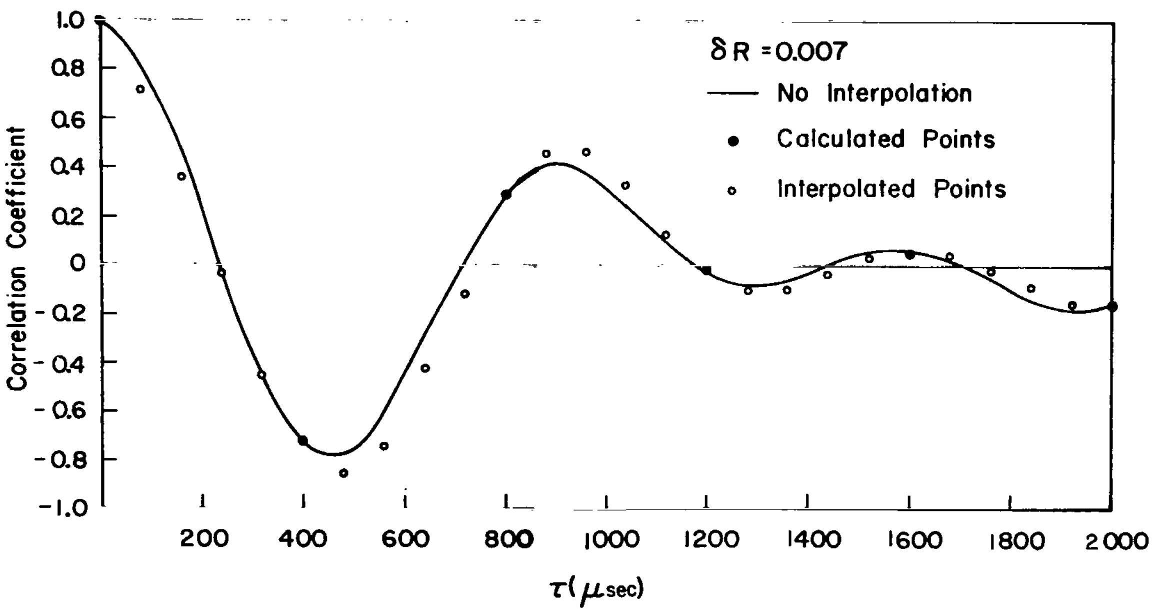


Figure 14. An interpolated correlation curve; δR = 0.007.

Overall Accuracy

In any data reduction method, irrespective of any special features it contains, the final question to be asked regards its overall accuracy. The test signal used here provides a surprisingly simple and revealing answer to this question. The signal was generated by passing white noise through a filter, the characteristics of which are provided by the manufacturer. Thus calculation of the covariance function of the test signal and its subsequent Fourier transformation will yield the spectrum of the test signal. This spectrum should then be directly comparable to the available transfer function of the filter used.

Such a comparison is shown in Figure 11. In spite of the stringency of the test created by the sharp edges of the filter function, a situation in which the frequency resolution offered by our own test and that of the manufacturer become of prime importance, a most acceptable agreement is observed.

CONCLUSIONS

The purpose of this report is to outline the special features and philosophy underlying the development of the piecewise correlation program. The concept of piecewise accumulation of mean values was developed initially to overcome storage limitations of general-purpose computers when they are used for statistical data reduction. As demonstrated in the section, Accumulation of Piecewise Averages, piecewise accumulation, combined with the use of appropriate recursion formulas, overcomes <u>all</u> limitations regarding the length of record which may be processed.

Furthermore, the work of the section, Derivatives of the Covariance Function, demonstrates how variations between such piecewise averages may be used to determine the probable accuracy of a result determined from a number of such pieces. Thus the statistical significance of a final result is immediately available. In fact, in the present program such statistical significance is used as a control parameter, the computation proceeding only until the degree of confidence in the result requested by the user is obtained. Also described is the way in which variation of these associated confidence intervals with record length processed may be used to detect the presence of nonstationary trends in that record. This feature is of particular importance in investigations of uncontrolled environments such as the atmosphere. The test results, presented in the section, Demonstration of the Piecewise Method, demonstrate that the calculated confidence intervals offer a very realistic estimate of the probable error of a computed result.

The facility to estimate derivatives of correlation functions is also included in the program. This is discussed in the section, Interpolation Between Lag Points, and confirmatory data are shown in the section, Demonstration of the Piecewise Method.

Finally, the overall accuracy of the scheme developed is shown in Figure 11.

In summary, the results of this work demonstrate that the concept of piecewise accumulation of mean values makes possible the processing of long statistical data records on general-purpose computers where storage limitations would otherwise be a problem. The associated error analysis is of value, in general, in optimizing computer time. However, the real power of this facility is to be found in the analysis of data records taken from uncontrolled environments, such as the atmosphere, where nonstationary trends might otherwise go undetected.

George C. Marshall Space Flight Center
National Aeronautics and Space Administration
Marshall Space Flight Center, Alabama 35812, March 1968
933-50-08-00-62

REFERENCES

1. Spiegel, M. R.: Theory and Problems of Statistics. Schaum Publishing Co., New York, 1961.

2. Bendat, J. S.; and Piersol, A. G.: Measurement and Analysis of Random Data. John Wiley and Sons, Inc., New York, 1966.

3. Jayroe, R. R.; and Su, M. Y.: Optimum Averaging Times of Meteorological Data with Time Dependent Means. AIAA Conference on Aerospace Meteorology, New Orleans, La., May 5-9, 1968.

NASA-Langley, 1970 —— 8 M531

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D. C. 20546

OFFICIAL BUSINESS

FIRST CLASS MAIL

POSTAGE AND FEES PAID
NATIONAL AERONAUTICS AND
SPACE ADMINISTRATION

06U 001 33 51 3DS 70329 00903
AIR FORCE WEAPONS LABORATORY /WLOL/
KIRTLAND AFB, NEW MEXICO 87117

ATT E. LOU BOWMAN, CHIEF,TECH. LIBRARY

POSTMASTER: If Undeliverable (Section 158 Postal Manual) Do Not Return

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

— NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Notes, and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546